

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: (Other than US) AMRAD OPERATIONS PTY LTD
(US only) Douglas James HILTON, Nicos Antony NICOLA, Alison FARLEY, Tracey WILLSON, Jian-Guo ZHANG, Warren ALEXANDER, Steven RAKAR, Louis FABRI, Tetsuo KOJIMA, Masatsugu MAEDA, Yasumfumi KIKUCHI, Andrew NASH

(ii) TITLE OF INVENTION: A NOVEL HAEMPOIETIN RECEPTOR AND GENETIC SEQUENCES ENCODING SAME

(iii) NUMBER OF SEQUENCES: 54

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSOR
(B) STREET: 400 GARDEN CITY PLAZA
(C) CITY: GARDEN CITY
(D) STATE: NEW YORK
(E) COUNTRY: UNITED STATES OF AMERICA
(F) ZIP: 11530-0299

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: CIP APPLICATION OF USSN 08/928,720
(B) FILING DATE: 10-MAR-1998

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PCT/GB97/02479
(B) FILING DATE: 11-SEP-1997

(viii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US APPLICATION NO. 08/928,720
(B) FILING DATE: 11-SEP-1997

(ix) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PO2246/96
(B) FILING DATE: 11-SEP-1996

(x) ATTORNEY/AGENT INFORMATION:

(A) NAME: DIGIGLIO, FRANK S
(B) REGISTRATION NO: 31,346
(C) REFERENCE/DOCKET NUMBER: 10857Z

(xi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: +516 742 4343
(B) TELEFAX: +516 742 4366

SUB
B

SEARCHED INDEXED
SERIALIZED FILED

- 68 -

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Trp Ser Xaa Trp Ser

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ACTCGCTCCA GATTCCCGCC TTTT

24

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TCCCCGCCTTT TTTCGACCCAT AGAT

24

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGTACTTGGC TTGGAAAGAGG AAAT

24

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- 69 -

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGGCTCACGT GCACGTCGGG TGGG

24

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGCTGCTGTT AAAGGGCTTC TC

22

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

(A/G)CTCCA(A/G)TC(A/G) CTCCA

15

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

(A/G)CTCCA(C/T)TC(A/G) CTCCA

15

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

- 70 -

AAGTGTGACC ATCATGTGGA C

21

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGAGGTGTTA AGGAGGCG

18

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGCCCGCGG GTCGCCCG

18

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1506 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1242

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGCACGAGCT TCGCTGTCCG CGCCCCAGTGA CGCGCGTGCG GACCCGAGCC CCAATCTGCA

-64

CCCCGCAGAC TCGCCCCCGC CCCATACCGG CGTTGCAGTC ACCGCCCGTT GCGCGCCACC

-4

CCC

-3

ATG CCC GCG GGT CGC CCG GGC CCC GTC GCC CAA TCC GCG CGG CCG CCG	48
Met Pro Ala Gly Arg Pro Gly. Pro Val Ala Gln Ser Ala Arg Arg Pro	
' 1 5 10 15	

CCG CGG CCG CTG TCC TCG CTG TGG TCG CCT CTG TTG CTC TGT GTC CTC	96
Pro Arg Pro Leu Ser S r Leu Trp Ser Pro Leu Leu Leu Cys Val Leu	
20 25 30	

GGG GTG CCT CGG GGC GGA TCG GGA GCC CAC ACA GCT GTA ATC AGC CCC Gly Val Pro Arg Gly Ser Gly Ala His Thr Ala Val Ile Ser Pro 35 40 45	144
CAG GAC CCC ACC CTT CTC ATC GGC TCC TCC CTG CAA GCT ACC TGC TCT Gln Asp Pro Thr Leu Leu Ile Gly Ser Ser Leu Gln Ala Thr Cys Ser 50 55 60	192
ATA CAT GGA GAC ACA CCT GGG GCC ACC GCT GAG GGG CTC TAC TGG ACC Ile His Gly Asp Thr Pro Gly Ala Thr Ala Glu Gly Leu Tyr Trp Thr 65 70 75 80	240
CTC AAT GGT CGC CGC CTG CCC TCT GAG CTG TCC CGC CTC CTT AAC ACC Leu Asn Gly Arg Arg Leu Pro Ser Glu Leu Ser Arg Leu Leu Asn Thr 85 90 95	288
TCC ACC CTG GCC CTG GCC CTG GCT AAC CTT AAT GGG TCC AGG CAG CAG Ser Thr Leu Ala Leu Ala Asn Leu Asn Gly Ser Arg Gln Gln 100 105 110	336
TCA GGA GAC AAT CTG GTG TGT CAC GCC CGA GAC GGC AGC ATT CTG GCT Ser Gly Asp Asn Leu Val Cys His Ala Arg Asp Gly Ser Ile Leu Ala 115 120 125	384
GGC TCC TGC CTC TAT GTT GGC TTG CCC CCT GAG AAG CCC TTT AAC ATC Gly Ser Cys Leu Tyr Val Gly Leu Pro Pro Glu Lys Pro Phe Asn Ile 130 135 140	432
AGC TGC TGG TCC CGG AAC ATG AAG GAT CTC ACG TGC CGC TGG ACA CCG Ser Cys Trp Ser Arg Asn Met Lys Asp Leu Thr Cys Arg Trp Thr Pro 145 150 155 160	480
GGT GCA CAC GGG GAG ACA TTC TTA CAT ACC AAC TAC TCC CTC AAG TAC Gly Ala His Gly Glu Thr Phe Leu His Thr Asn Tyr Ser Leu Lys Tyr 165 170 175	528
AAG CTG AGG TGG TAC GGT CAG GAT AAC ACA TGT GAG GAG TAC CAC ACT Lys Leu Arg Trp Tyr Gly Gln Asp Asn Thr Cys Glu Glu Tyr His Thr 180 185 190	576
G TG GGC CCT CAC TCA TGC CAT ATC CCC AAG GAC CTG GCC CTC TTC ACT Val Gly Pro His Ser Cys His Ile Pro Lys Asp Leu Ala Leu Phe Thr 195 200 205	624
CCC TAT GAG ATC TGG GTG GAA GCC ACC AAT CGC CTA GGC TCA GCA AGA Pro Tyr Glu Ile Trp Val Glu Ala Thr Asn Arg Leu Gly Ser Ala Arg 210 215 220	672
TCT GAT GTC CTC ACA CTG GAT GTC CTG GAC GTG GTG ACC ACC GAC CCC Ser Asp Val Leu Thr Leu Asp Val Leu Asp Val Val Thr Thr Asp Pro 225 230 235 240	720
CCA CCC GAC GTG CAC GTG AGC CGC GTT GGG GGC CTG GAG GAC CAG CTG Pro Pro Asp Val His Val Ser Arg Val Gly Leu Glu Asp Gln Leu 245 250 255	768
AGT GTG CGC TGG GTC TCA CCA CCA GCT CTC AAG GAT TTC CTC TTC CAA Ser Val Arg Trp Val Ser Pro Pro Ala Leu Lys Asp Phe Leu Phe Gln 260 265 270	816
GCC AAG TAC CAG ATC CGC TAC CGC GTG GAG GAC AGC GTG GAC TGG AAG Ala Lys Tyr Gln Ile Arg Tyr Arg Val Glu Asp Ser Val Asp Trp Lys 275 280 285	864
G TG GTG GAT GAC GTC AGC AAC CAG ACC TCC TGC CGT CTC GCG GGC CTG Val Val Asp Asp Val Ser Asn Gln Thr Ser Cys Arg Leu Ala Gly Leu 290 295 300	912
AAG CCC GGC ACC GTT TAC TTC GTC CAA GTG CGT TGT AAC CCA TTC GGG Lys Pro Gly Thr Val Tyr Phe Val Gln Val Arg Cys Asn Pro Phe Gly 305 310 315 320	960
ATC TAT GGG TCG AAA AAG GCG GGA ATC TGG AGC GAG TGG AGC CAC CCC Ile Tyr Gly Ser Lys Lys Ala Gly Ile Trp Ser Glu Trp Ser His Pro	1008

- 72 -

325	330	335	
ACC GCT GCC TCC ACC CCT CGA AGT GAG CGC CCG GGC CCG GGC GGC GGG Thr Ala Ala Ser Thr Pro Arg Ser Glu Arg Pro Gly Pro Gly Gly Gly	340	345	1056
GTG TGC GAG CCG CGG GGC GGC GAG CCC AGC TCG GGC CCG GTG CGG CGC Val Cys Glu Pro Arg Gly Gly Glu Pro Ser Ser Gly Pro Val Arg Arg	355	360	1104
GAG CTC AAG CAG TTC CTC GGC TCG CTC AAG AAG CAC GCA TAC TGC TCG Glu Leu Lys Gln Phe Leu Gly Trp Leu Lys Lys His Ala Tyr Cys Ser	370	375	1152
AAC CTT AGT TTC CGC CTG TAC GAC CAG TGG CGT GCT TGG ATG CAG AAG Asn Leu Ser Phe Arg Leu Tyr Asp Gln Trp Arg Ala Trp Met Gln Lys	385	390	1200
TCA CAC AAG ACC CGA AAC CAG GTC CTG CCG GCT AAA CTC TAAGGATAGG Ser His Lys Thr Arg Asn Gln Val Leu Pro Ala Lys Leu	405	410	1249
CCATCCTCCT GCTGGTCAG ACCTGGAGGC TCACCTGAAT TGGAGCCCT CTGTACCATC TGGGCAACAA AGAACCTAC CAGAGGCTGG GGCACAATGA GCTCCCACAA CCACAGCTTT			1309
GGTCCACATG ATGGTCACAC TTGGATATAAC CCCAGTGTGG GTAAAGTTGG GGTATTGCAG GGCCTCCCAA CAATCTCTTT AAATAAATAA AGGAGTTGTT CAGGTAAAAA AAAAAAAAAA			1369
AAAAAAAAAAA AAAAAAAA			1429
			1489
			1506

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 413 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Pro Ala Gly Arg Pro Gly Pro Val Ala Gln Ser Ala Arg Arg Pro			
1	5	10	15
Pro Arg Pro Leu Ser Ser Leu Trp Ser Pro Leu Leu Cys Val Leu			
20	25	30	
Gly Val Pro Arg Gly Gly Ser Gly Ala His Thr Ala Val Ile Ser Pro			
35	40	45	
Gln Asp Pro Thr Leu Leu Ile Gly Ser Ser Leu Gln Ala Thr Cys Ser			
50	55	60	
Ile His Gly Asp Thr Pro Gly Ala Thr Ala Glu Gly Leu Tyr Trp Thr			
65	70	75	80
Leu Asn Gly Arg Arg Leu Pro Ser Glu Leu Ser Arg Leu Leu Asn Thr			
85	90	95	
Ser Thr Leu Ala Leu Ala Leu Asn Leu Asn Gly Ser Arg Gln Gln			
100	105	110	
Ser Gly Asp Asn Leu Val Cys His Ala Arg Asp Gly Ser Ile Leu Ala			
115	120	125	
Gly Ser Cys Leu Tyr Val Gly Leu Pro Pro Glu Lys Pro Phe Asn Ile			
130	135	140	
Ser Cys Trp Ser Arg Asn Met Lys Asp Leu Thr Cys Arg Trp Thr Pro			
145	150	155	160

- 73 -

Gly Ala His Gly Glu Thr Phe Leu His Thr Asn Tyr Ser Leu Lys Tyr
 165 170 175
 Lys Leu Arg Trp Tyr Gly Gln Asp Asn Thr Cys Glu Glu Tyr His Thr
 180 185 190
 Val Gly Pro His Ser Cys His Ile Pro Lys Asp Leu Ala Leu Phe Thr
 195 200 205
 Pro Tyr Glu Ile Trp Val Glu Ala Thr Asn Arg Leu Gly Ser Ala Arg
 210 215 220
 Ser Asp Val Leu Thr Leu Asp Val Leu Asp Val Val Thr Thr Asp Pro
 225 230 235 240
 Pro Pro Asp Val His Val Ser Arg Val Gly Gly Leu Glu Asp Gln Leu
 245 250 255
 Ser Val Arg Trp Val Ser Pro Pro Ala Leu Lys Asp Phe Leu Phe Gln
 260 265 270
 Ala Lys Tyr Gln Ile Arg Tyr Arg Val Glu Asp Ser Val Asp Trp Lys
 275 280
 Val Val Asp Asp Val Ser Asn Gln Thr Ser Cys Arg Leu Ala Gly Leu
 290 295 300
 Lys Pro Gly Thr Val Tyr Phe Val Gln Val Arg Cys Asn Pro Phe Gly
 305 310 315 320
 Ile Tyr Gly Ser Lys Lys Ala Gly Ile Trp Ser Glu Trp Ser His Pro
 325 330 335
 Thr Ala Ala Ser Thr Pro Arg Ser Glu Arg Pro Gly Pro Gly Gly
 340 345 350
 Val Cys Glu Pro Arg Gly Gly Glu Pro Ser Ser Gly Pro Val Arg Arg
 355 360 365
 Glu Leu Lys Gln Phe Leu Gly Trp Leu Lys Lys His Ala Tyr Cys Ser
 370 375 380
 Asn Leu Ser Phe Arg Leu Tyr Asp Gln Trp Arg Ala Trp Met Gln Lys
 385 390 395 400
 Ser His Lys Thr Arg Asn Gln Val Leu Pro Ala Lys Leu
 405 410

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1549 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..1278

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGCACGAGCT TCGCTGTCCG CGCCCCAGTGA CGCGCGTGCG GACCCGAGCC CCAATCTGCA
 CCCCCGCAGAC TCGCCCCCGC CCCATACCGG CGTTGCCAGTC ACCGCCCCGTT GCGCGCCACC

-65

-5

CCCCA

-1

ATG CCC GCG GGT CGC CCG GGCG CCC GTC GCC CAA TCC GCG CGG CGG CCG Met Pro Ala Gly Arg Pro Gly Pro Val Ala Gln Ser Ala Arg Arg Pro 1 5 10 15	48
CCG CGG CCG CTG TCC TCG CTG TGG TCG CCT CTG TTG CTC TGT GTC CTC Pro Arg Pro Leu Ser Ser Leu Trp Ser Pro Leu Leu Leu Cys Val Leu 20 25 30	96
GGG GTG CCT CGG GGC GGA TCG GGA GCC CAC ACA GCT GTA ATC AGC CCC Gly Val Pro Arg Gly Gly Ser Gly Ala His Thr Ala Val Ile Ser Pro 35 40 45	144
CAG GAC CCC ACC CTT CTC ATC GGC TCC TCC CTG CAA GCT ACC TGC TCT Gln Asp Pro Thr Leu Leu Ile Gly Ser Ser Leu Gln Ala Thr Cys Ser 50 55 60	192
ATA CAT GGA GAC ACA CCT GGG GCC ACC GCT GAG GGG CTC TAC TGG ACC Ile His Gly Asp Thr Pro Gly Ala Thr Ala Glu Gly Leu Tyr Trp Thr 65 70 75 80	240
CTC AAT GGT CGC CGC CTG CCC TCT GAG CTG TCC CGC CTC CTT AAC ACC Leu Asn Gly Arg Arg Leu Pro Ser Glu Leu Ser Arg Leu Leu Asn Thr 85 90 95	288
TCC ACC CTG GCC CTG GCC CTG GCT AAC CTT AAT GGG TCC AGG CAG CAG Ser Thr Leu Ala Leu Ala Asn Leu Asn Gly Ser Arg Gln Gln 100 105 110	336
TCA GGA GAC AAT CTG GTG TGT CAC GCC CGA GAC GGC AGC ATT CTG GCT Ser Gly Asp Asn Leu Val Cys His Ala Arg Asp Gly Ser Ile Leu Ala 115 120 125	384
GCC TCC TGC CTC TAT GTT GGC TTG CCC CCT GAG AAG CCC TTT AAC ATC Gly Ser Cys Leu Tyr Val Gly Leu Pro Pro Glu Lys Pro Phe Asn Ile 130 135 140	432
AGC TGC TGG TCC CGG AAC ATG AAG GAT CTC ACG TGC CGC TGG ACA CCG Ser Cys Trp Ser Arg Asn Met Lys Asp Leu Thr Cys Arg Trp Thr Pro 145 150 155 160	480
GGT GCA CAC CGG GAG ACA TTC TTA CAT ACC AAC TAC TCC CTC AAG TAC Gly Ala His Gly Glu Thr Phe Leu His Thr Asn Tyr Ser Leu Lys Tyr 165 170 175	528
AAG CTG AGG TCG TAC GGT CAG GAT AAC ACA TGT GAG GAG TAC CAC ACT Lys Leu Arg Trp Tyr Gly Gln Asp Asn Thr Cys Glu Glu Tyr His Thr 180 185 190	576
GTG GGC CCT CAC TCA TGC CAT ATC CCC AAG GAC CTG GCC CTC TTC ACT Val Gly Pro His Ser Cys His Ile Pro Lys Asp Leu Ala Leu Phe Thr 195 200 205	624
CCC TAT GAG ATC TGG GTG GAA GCC ACC AAT CGC CTA CGC TCA GCA AGA Pro Tyr Glu Ile Trp Val Glu Ala Thr Asn Arg Leu Gly Ser Ala Arg 210 215 220	672
TCT GAT GTC CTC ACA CTG GAT GTC CTG GAC GTG GTG ACC ACG GAC CCC Ser Asp Val Leu Thr Leu Asp Val Leu Asp Val Val Thr Thr Asp Pro 225 230 235 240	720
CCA CCC GAC CTG CAC GTG AGC CGC GTT GGG GGC CTG GAG GAC CAG CTG Pro Pro Asp Val His Val Ser Arg Val Gly Gly Leu Glu Asp Gln Leu 245 250 255	768
AGT GTG CGC TGG GTC TCA CCA CCA GCT CTC AAG GAT TTC CTC TTC CAA Ser Val Arg Trp Val Ser Pro Pro Ala Leu Lys Asp Phe Leu Phe Gln 260 265 270	816
GCC AAG TAC CAG ATC CGC TAC CGC GTG GAG GAC AGC GTG GAC TGG AAG Ala Lys Tyr Gln Ile Arg Tyr Arg Val Glu Asp Ser Val Asp Trp Lys 275 280 285	864

G TG GAT GAC GTC AGC AAC CAG ACC TCC TGC CGT CTC GCG GGC CTG Val Val Asp Asp Val Ser Asn Gln Thr Ser Cys Arg Leu Ala Gly Leu 290 295 300	912
AAG CCC GGC ACC GTT TAC TTC GTC CAA GTG CGT TGT AAC CCA TTC GGG Lys Pro Gly Thr Val Tyr Phe Val Gln Val Arg Cys Asn Pro Phe Gly 305 310 315 320	960
ATC TAT GGG TCG AAA AAG GCG GGA ATC TGG AGC GAG TGG AGC CAC CCC Ile Tyr Gly Ser Lys Lys Ala Gly Ile Trp Ser Glu Trp Ser His Pro 325 330 335	1008
ACC GCT GCC TCC ACC CCT CGA AGT GAG CGC CCG GGC CCG GGC GGC GGG Thr Ala Ala Ser Thr Pro Arg Ser Glu Arg Pro Gly Pro Gly Gly 340 345 350	1056
G TG GAG CCG CGG GGC GGC GAG CCC AGC TCG GGC CCG GTG CGG CCC Val Cys Glu Pro Arg Gly Glu Pro Ser Ser Gly Pro Val Arg Arg 355 360 365	1104
GAG CTC AAG CAG TTC CTC CGC TGG CTC AAG AAG CAC GCA TAC TGC TCG Glu Leu Lys Gln Phe Leu Gly Trp Leu Lys Lys His Ala Tyr Cys Ser 370 375 380	1152
AAC CTT AGT TTC CGC CTG TAC GAC CAG TGG CGT GCT TGG ATG CAG AAG Asn Leu Ser Phe Arg Leu Tyr Asp Gln Trp Arg Ala Trp Met Gln Lys 385 390 395 400	1200
TCA CAC AAG ACC CGA AAC CAG GAC GAG GGG ATC CTG CCT TCG GGC AGA Ser His Lys Thr Arg Asn Gln Asp Glu Gly Ile Leu Pro Ser Gly Arg 405 410 415	1248
CGG GGT GCG GCG AGA GGT CCT GCC GGT TAAACTCTAA GGATAGGCCA Arg Gly Ala Ala Arg Gly Pro Ala Gly 420 425	1295
TCCTCCTGCT GGGTCAGACC TGGAGGCTCA CCTGAATTGG ACCCCCTCTG TACCATCTGG GCAACAAAGA AACCTACCAAG AGGCTGGGGC ACAATGAGCT CCCACAACCA CAGCTTTGGT CCACATGATG GTCACACTTG GATATAACCCC AGTGTGGTA AGGTTGGGT ATTCCAGGGC CTCCCAACAA TCTCTTTAAA TAAATAAAGG AGTTGTTCAAG GTAAAAAAAAA AAAAAAAA AAAAAAA AAAA	1355 1415 1475 1535 1549

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 425 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Pro Ala Gly Arg Pro Gly Pro Val Ala Gln Ser Ala Arg Arg Pro 1 5 10 15
Pro Arg Pro Leu Ser Ser Leu Trp Ser Pro Leu Leu Leu Cys Val Leu 20 25 30
Gly Val Pro Arg Gly Gly Ser Gly Ala His Thr Ala Val Ile Ser Pro 35 40 45
Gln Asp Pro Thr Leu Leu Ile Gly Ser Ser Leu Gln Ala Thr Cys Ser 50 55 60
Ile His Gly Asp Thr Pro Gly Ala Thr Ala Glu Gly Leu Tyr Trp Thr 65 70 75 80

SEQUENCE DEDUCED

- 76 -

Leu	Asn	Gly	Arg	Arg	Leu	Pro	Ser	Glu	Leu	Ser	Arg	Leu	Leu	Asn	Thr
85								90						95	
Ser	Thr	Leu	Ala	Leu	Ala	Leu	Ala	Asn	Leu	Asn	Gly	Ser	Arg	Gln	Gln
100								105					110		
Ser	Gly	Asp	Asn	Leu	Val	Cys	His	Ala	Arg	Asp	Gly	Ser	Ile	Leu	Ala
115						120							125		
Gly	Ser	Cys	Leu	Tyr	Val	Gly	Leu	Pro	Pro	Glu	Lys	Pro	Phe	Asn	Ile
130						135				140					
Ser	Cys	Trp	Ser	Arg	Asn	Met	Lys	Asp	Leu	Thr	Cys	Arg	Trp	Thr	Pro
145						150				155			160		
Gly	Ala	His	Gly	Glu	Thr	Phe	Leu	His	Thr	Asn	Tyr	Ser	Leu	Lys	Tyr
165								170					175		
Lys	Leu	Arg	Trp	Tyr	Gly	Gln	Asp	Asn	Thr	Cys	Glu	Glu	Tyr	His	Thr
180						185				190					
Val	Gly	Pro	His	Ser	Cys	His	Ile	Pro	Lys	Asp	Leu	Ala	Leu	Phe	Thr
195							200					205			
Pro	Tyr	Glu	Ile	Trp	Val	Glu	Ala	Thr	Asn	Arg	Leu	Gly	Ser	Ala	Arg
210						215				220					
Ser	Asp	Val	Leu	Thr	Leu	Asp	Val	Leu	Asp	Val	Val	Thr	Thr	Asp	Pro
225						230				235			240		
Pro	Pro	Asp	Val	His	Val	Ser	Arg	Val	Gly	Gly	Leu	Glu	Asp	Gln	Leu
245								250					255		
Ser	Val	Arg	Trp	Val	Ser	Pro	Pro	Ala	Leu	Lys	Asp	Phe	Leu	Phe	Gln
260								265					270		
Ala	Lys	Tyr	Gln	Ile	Arg	Tyr	Arg	Val	Glu	Asp	Ser	Val	Asp	Trp	Lys
275								280					285		
Val	Val	Asp	Asp	Val	Ser	Asn	Gln	Thr	Ser	Cys	Arg	Leu	Ala	Gly	Leu
290								295					300		
Lys	Pro	Gly	Thr	Val	Tyr	Phe	Val	Gln	Val	Arg	Cys	Asn	Pro	Phe	Gly
305								310			315			320	
Ile	Tyr	Gly	Ser	Lys	Lys	Ala	Gly	Ile	Trp	Ser	Glu	Trp	Ser	His	Pro
325									330					335	
Thr	Ala	Ala	Ser	Thr	Pro	Arg	Ser	Glu	Arg	Pro	Gly	Pro	Gly	Gly	
340								345					350		
Val	Cys	Glu	Pro	Arg	Gly	Gly	Glu	Pro	Ser	Ser	Gly	Pro	Val	Arg	Arg
355								360					365		
Glu	Leu	Lys	Gln	Phe	Leu	Gly	Trp	Leu	Lys	Lys	His	Ala	Tyr	Cys	Ser
370						375						380			
Asn	Leu	Ser	Phe	Arg	Leu	Tyr	Asp	Gln	Trp	Arg	Ala	Trp	Met	Gln	Lys
385								390				395			400
Ser	His	Lys	Thr	Arg	Asn	Gln	Asp	Glu	Gly	Ile	Leu	Pro	Ser	Gly	Arg
405									410					415	
Arg	Gly	Ala	Ala	Arg	Gly	Pro	Ala	Gly							
420															425

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 938 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

- 77 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..468

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGC ACC GTT TAC TTC GTC CAA GTG CGT TGT AAC CCA TTC GGG ATC TAT 1 5	Gly Thr Val Tyr Phe Val Gln Val Arg Cys Asn Pro Phe Gly Ile Tyr 10 15	48
GGG TCG AAA AAG GCG GGA ATC TGG AGC GAG TGG AGC CAC CCC ACC GCT 20 25	Gly Ser Lys Lys Ala Gly Ile Trp Ser Glu Trp Ser His Pro Thr Ala 30	96
GCC TCC ACC CCT CGA AGT GAG CGC CCG GGC CCG GGC GGC GGG GTG TGC 35 40	Ala Ser Thr Pro Arg Ser Glu Arg Pro Gly Pro Gly Gly Val Cys 45	144
GAG CCG CGG GGC GGC GAG CCC AGC TCG GGC CCG GTG CGG CGC GAG CTC 50 55	Glu Pro Arg Gly Gly Glu Pro Ser Ser Gly Pro Val Arg Arg Glu Leu 60	192
AAG CAG TTC CTC GGC TGG CTC AAG AAG CAC GCA TAC TGC TCG AAC CTT 65 70	Lys Gln Phe Leu Gly Trp Leu Lys Lys His Ala Tyr Cys Ser Asn Leu 75 80	240
AGT TTC CGC CTG TAC GAC CAG TGG CGT GCT TGG ATG CAG AAG TCA CAC 85 90	Ser Phe Arg Leu Tyr Asp Gln Trp Arg Ala Trp Met Gln Lys Ser His 95	288
AAG ACC CGA AAC CAG GTA GGA AAG TTG GGG GAG GCT TGC GTG GGG GGT 100 105	Lys Thr Arg Asn Gln Val Gly Lys Leu Gly Glu Ala Cys Val Gly Gly 110	336
AAA CGA GCA GAG GAA GAG AGA GAC CCG GGT GAG CAG CCT CCA CAA CAC 115 120	Lys Gly Ala Glu Glu Glu Arg Asp Pro Gly Glu Gln Pro Pro Gln His 125	384
CGC ACT CTT CTT TCC AAG CAC AGG ACG AGG GGA TCC TGC CCT CGG GCA 130 135	Arg Thr Leu Leu Ser Lys His Arg Thr Arg Gly Ser Cys Pro Arg Ala 140	432
GAC CGG GTG CGG CGA GAG GTA AGG CGG TCT CGG TGAGTGGGGC CTACAGCAGT 145 150	Asp Gly Val Arg Arg Glu Val Arg Gly Ser Gly 155	485
CTAGATGAGG CCCTTTCCCCC TCCTTCGGTG TTGCTCAAAG CGATCTCTTA GTGCTCATTT		545
CACCCACTGC AAAGAGCCCC AGGTTTTACT GCATCATCAA GTGCTGAAG GGTCCAGGCT		605
TAATGTGGCC TCTTTCTGC CCTCAGGTCC TGCCGGCTAA ACTCTAAGGA TAGGCCATCC		665
TCCTGCTGGG TCAGACCTGG AGGCTCACCT GAATTGGAGC CCTCTGTAC CTATCTGGGC		725
AAACAAAGAAA CCTACCATGA GGCTGGGGCA CAATGAGCTC CCACAACCAC AGCTTTGGTC		785
CACATGATGG TCACACTTGG ATATAACCCA GTGTGGTAA GGTTGGGGTA TTGCAGGGCC		845
TCCCAACAAT CTCTTTAAAT AAATAAAGGA GTTGTTCAGG TAAAAAAA AAAAAAAA		905
AAAAAAAAAA AAAAAAAA AAAAAAAA AAA		938

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 155 amino acids
- (B) TYPE: amino acid

- 78 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Gly	Thr	Val	Tyr	Phe	Val	Gln	Val	Arg	Cys	Asn	Pro	Phe	Gly	Ile	Tyr	
1					5				10					15		
Gly	Ser	Lys	Lys	Ala	Gly	Ile	Trp	Ser	Glu	Trp	Ser	His	Pro	Thr	Ala	
					20			25			30					
Ala	Ser	Thr	Pro	Arg	Ser	Glu	Arg	Pro	Gly	Pro	Gly	Gly	Val	Cys		
					35		40				45					
Glu	Pro	Arg	Gly	Gly	Glu	Pro	Ser	Ser	Gly	Pro	Val	Arg	Arg	Glu	Leu	
					50		55				60					
Lys	Gln	Phe	Leu	Gly	Trp	Leu	Lys	Lys	His	Ala	Tyr	Cys	Ser	Asn	Leu	
					65		70		75		80					
Ser	Phe	Arg	Leu	Tyr	Asp	Gln	Trp	Arg	Ala	Trp	Met	Gln	Lys	Ser	His	
					85		90		95							
Lys	Thr	Arg	Asn	Gln	Val	Gly	Lys	Leu	Gly	Glu	Ala	Cys	Val	Gly	Gly	
					100		105				110					
Lys	Gly	Ala	Glu	Glu	Glu	Arg	Asp	Pro	Gly	Glu	Gln	Pro	Pro	Gln	His	
					115		120				125					
Arg	Thr	Leu	Leu	Ser	Lys	His	Arg	Thr	Arg	Gly	Ser	Cys	Pro	Arg	Ala	
					130		135				140					
Asp	Gly	Val	Arg	Arg	Glu	Val	Arg	Gly	Ser	Gly						
					145		150				155					

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 834 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..834

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCC	ACC	CTT	CTC	ATC	GTC	TCC	CTG	CAA	GCT	ACC	TGC	TCT	ATA	CAT		98	
Pro																	
51					55				60					65			
GGA	GAC	ACA	CCT	GGG	GCC	ACC	GCT	GAG	GGG	CTC	TAC	TGG	ACC	CTC	AAT		146
Gly	Asp	Thr	Pro	Gly	Ala	Thr	Ala	Glu	Gly	Leu	Tyr	Trp	Thr	Leu	Asn		
					70			75									
GGT	CGC	CGC	CTG	CCC	TCT	GAG	CTG	TCC	CGC	CTC	CTT	AAC	ACC	TCC	ACC		194
Gly	Arg	Arg	Leu	Pro	Ser	Glu	Leu	Ser	Arg	Leu	Leu	Asn	Thr	Ser	Thr		
					85			90				95					
CTG	GCC	CTG	GCC	CTG	GCT	AAC	CTT	AAT	GGG	TCC	AGG	CAG	CAG	TCA	GGA		242
Leu	Ala	Leu	Ala	Leu	Asn	Leu	Asn	Gly	Ser	Arg	Gln	Gln	Ser	Gly			
					100		105				110						
GAC	AAT	CTG	GTG	TGT	CAC	GCC	CGA	GAC	GCC	AGC	ATT	CTG	GCT	GCC	TCC		290

- 79 -

Asp Asn Leu Val Cys His Ala Arg Asp Gly Ser Ile Leu Ala Gly Ser		
115 120 125 130		
TGC CTC TAT GTT GGC TTG CCC CCT GAG AAG CCC TTT AAC ATC AGC TGC	338	
Cys Leu Tyr Val Gly Leu Pro Pro Glu Lys Pro Phe Asn Ile Ser Cys		
135 140 145		
TGG TCC CGG AAC ATG AAG GAT CTC ACG TGC CGC TGG ACA CCG GGT GCA	386	
Trp Ser Arg Asn Met Lys Asp Leu Thr Cys Arg Trp Thr Pro Gly Ala		
150 155 200		
CAC CGG GAG ACA TTC TTA CAT ACC AAC TAC TCC CTC AAG TAC AAG CTG	434	
His Gly Glu Thr Phe Leu His Thr Asn Tyr Ser Leu Lys Tyr Lys Leu		
205 210 215		
AGG TGG TAC GGT CAG GAT AAC ACA TGT GAG GAG TAC CAC ACT GTG GGG	482	
Arg Trp Tyr Gly Gln Asp Asn Thr Cys Glu Glu Tyr His Thr Val Gly		
220 225 230		
CCC CAC TCA TGC CAT ATC CCC AAG GAC CTG GCC CTC TTC ACT CCC TAT	530	
Pro His Ser Cys His Ile Pro Lys Asp Leu Ala Leu Phe Thr Pro Tyr		
235 240 245 250		
GAG ATC TGG GTG GAA GCC ACC AAT CCC CTA GGC TCA GCA AGA TCT GAT	578	
Glu Ile Trp Val Glu Ala Thr Asn Arg Leu Gly Ser Ala Arg Ser Asp		
255 260 265		
GTC CTC ACA CTG GAT GTC CTG GAC GTG GTG ACC ACG GAC CCC CCA CCC	626	
Val Leu Thr Leu Asp Val Leu Asp Val Val Thr Thr Asp Pro Pro Pro		
270 275 280		
GAC GTG CAC GTG AGC CGC GTT GGG GGC CTG GAG GAC CAG CTG AGT GTG	674	
Asp Val His Val Ser Arg Val Gly Gly Leu Glu Asp Gln Leu Ser Val		
285 290 295		
CGC TGG GTC TCA CCA CCA CCT CTC AAG GAT TTC CTC TTC CAA GCC AAG	722	
Arg Trp Val Ser Pro Pro Ala Leu Lys Asp Phe Leu Phe Gln Ala Lys		
300 305 310		
TAC CAG ATC CGC TAC CGC GTG GAG GAC AGC GTG GAC TGG AAG GTG GTG	770	
Tyr Gln Ile Arg Tyr Arg Val Glu Asp Ser Val Asp Trp Lys Val Val		
315 320 325 330		
GAT GAC GTC AGC AAC CAG ACC TCC TGC CGT CTC GCG GGC CTG AAG CCC	818	
Asp Asp Val Ser Asn Gln Thr Ser Cys Arg Leu Ala Gly Leu Lys Pro		
335 340 345		
GCC ACC GTT TAC TTC GTC CAA GTG CGT TGT AAC CCA TTC GGG ATC TAT	866	
Gly Thr Val Tyr Phe Val Gln Val Arg Cys Asn Pro Phe Gly Ile Tyr		
350 355 360		
GGG TCG AAA AAG GCG GGA	894	
Gly Ser Lys Lys Ala Gly		
365		

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 278 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Pro Thr Leu Leu Ile Gly Ser Ser Leu Gln Ala Thr Cys Ser Ile His
 51 55 60 65

Gly Asp Thr Pro Gly Ala Thr Ala Glu Gly Leu Tyr Trp Thr Leu Asn
 70 75 80

- 80 -

Gly Arg Arg Leu Pro Ser Glu Leu Ser Arg Leu Leu Asn Thr Ser Thr
 85 90 95
 Leu Ala Leu Ala Leu Ala Asn Leu Asn Gly Ser Arg Gln Gln Ser Gly
 100 105 110
 Asp Asn Leu Val Cys His Ala Arg Asp Gly Ser Ile Leu Ala Gly Ser
 115 120 125 130
 Cys Leu Tyr Val Gly Leu Pro Pro Glu Lys Pro Phe Asn Ile Ser Cys
 135 140 145
 Trp Ser Arg Asn Met Lys Asp Leu Thr Cys Arg Trp Thr Pro Gly Ala
 150 155 200
 His Gly Glu Thr Phe Leu His Thr Asn Tyr Ser Leu Lys Tyr Lys Leu
 205 210 215
 Arg Trp Tyr Gly Gln Asp Asn Thr Cys Glu Glu Tyr His Thr Val Gly
 220 225 230
 Pro His Ser Cys His Ile Pro Lys Asp Leu Ala Leu Phe Thr Pro Tyr
 235 240 245 250
 Glu Ile Trp Val Glu Ala Thr Asn Arg Leu Gly Ser Ala Arg Ser Asp
 255 260 265
 Val Leu Thr Leu Asp Val Leu Asp Val Val Thr Thr Asp Pro Pro Pro
 270 275 280
 Asp Val His Val Ser Arg Val Gly Gly Leu Glu Asp Gln Leu Ser Val
 285 290 295
 Arg Trp Val Ser Pro Pro Ala Leu Lys Asp Phe Leu Phe Gln Ala Lys
 300 305 310
 Tyr Gln Ile Arg Tyr Arg Val Glu Asp Ser Val Asp Trp Lys Val Val
 315 320 325 330
 Asp Asp Val Ser Asn Gln Thr Ser Cys Arg Leu Ala Gly Leu Lys Pro
 335 340 345
 Gly Thr Val Tyr Phe Val Gln Val Arg Cys Asn Pro Phe Gly Ile Tyr
 350 355 360
 Gly Ser Lys Lys Ala Gly
 365

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 143 base pairs
- (B) TYPE: nucleic acids
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGCATGAAGG CTTAGGGTGG GGATCGGTAG GACCCATGCA CCCAGAGAAA CGGACTGGTG	60
GCAACTTCA AACTCTCTGG GGAAGGAAGA AGGGCTGAAA GAGG	104
ATG AAC GGG CTC AGA CAC AGC TGT AAT CAG CCC CCA GGA Met Asn Gly Leu Arg His Ser Cys Asn Gln Pro Pro Gly	143 5 10

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids

- 81 -

(B) TYPE: amino acids
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Met Asn Gly Leu Arg His Ser Cys Asn Gln Pro Pro Gly
 5 10

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1930 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGCACGAGCT TCGCTGTCCG CGCCCAGTGA CGCGCGTGCG GACCCGAGCC CCAATCTGCA	60
CCCCCGCAGAC TCGCCCCCGC CCCATACCGG CGTTGCAGTC ACCGCCCGTT GCGCGCCACC	120
CCCAATGCCG CGGGGTGCGC CGGGCCCCGT CGCCCAATCC GCGCGGCGGC CGCCGCGGCC	180
GCTGTCCCTCG CTGTGGTCGC CTCTGTTGCT CTGTGTCCCTC GGGGTGCCTC GGGGGGGATC	240
GGGAGCCAC ACAGCTGTAA TCAGCCCCCA GGACCCCACC CTTCTCATCG GCTCCTCCCT	300
GCAAGCTACC TGCTCTATAAC ATGGAGACAC ACCTGGGGCC ACCGCTGAGG GGCTCTACTG	360
GACCCCTCAAT GGTGCGGCC TGCCCTCTGA GCTGTCCCGC CTCCCTAACCA CCTCCACCCCT	420
GGCCCTGGCC CTGGCTAACCT TTAATGGGTC CAGGCAGCAG TCAGGAGACA ATCTGGTGTG	480
TCACGCCCCA GACGGCAGCA TTCTGGCTGG CTCCTGCCCTC TATGTTGGCT TGCCCCCTGA	540
GAAGCCCCTTT AACATCAGCT GCTGGTCCCCG GAACATGAAG GATCTCACGT GCCGCTGGAC	600
ACCGGGTGCA CACGGGGAGA CATTCTTACA TACCAACTAC TCCCTCAAGT ACAAGCTGAG	660
GTGGTACGGT CAGGATAACA CATGTGAGGA GTACCACACT GTGGGCCCTC ACTCATGCCA	720
TATCCCCAAG GACCTGGCCC TCTTCACTCC CTATGAGATC TGGGTGGAAG CCACCAATCG	780
CCTAGGCTCA GCAAGATCTG ATGTCTCAC ACTGGATGTC CTGGACGTGG TGACCACGGA	840
CCCCCCACCC GACGTGCACG TGAGCCGCGT TGGGGCCTG GAGGACCAGC TGAGTGTGCG	900
CTGGGTCTCA CCACCAAGCTC TCAAGGATTTC CCTCTTCAA GCCAAGTACC AGATCCGCTA	960
CCGCGTGGAG GACAGCGTGG ACTGGAAGGT GGTGGATGAC GTCAAGCAACC AGACCTCCTG	1020
CCGTCTCGCG GGCCTGAAGC CCGGCACCGT TTACTTCGTC CAAGTGCCTT GTAACCCATT	1080
CGGGATCTAT GGGTCGAAAA AGGGCGGAAT CTGGAGCGAG TGGAGCCACC CCACCGCTGC	1140
CTCCACCCCT CGAAGTGAGC GCCCCGGGCC GGGCGGCGGG GTGTGCGAGC CGCGGGCGG	1200
CGAGCCCCAGC TCGGGCCCGG TGCGGCGCGA GCTCAAGCAG TTCTCGGCT GGCTCAAGAA	1260
GCACGCATAC TGCTCGAACCT TAGTTTCGG CCTGTACGAC CAGTGGCGTG CTTGGATGCA	1320
GAAGTCACAC AAGACCCGAA ACCAGGTAGG AAAGTTGGGG GAGGCTTGCG TGGGGGGTAA	1380
AGGAGCAGAG GAAGAGAGAG ACCCGGGGTGA GCAGCCTCCA CAACACCGCA CTCTTCTTTC	1440

CAAGCACAGG ACGAGGGGAT CCTGCCCTCG GGCAGACGGG GTGCGGGAG AGGTAAGGGG	1500
GTCTGGGTGA GTGGGGCCTA CAGCAGTCTA GATGAGGCC CTTTCCCCCTCC TTCTGGTGTG	1560
CTCAAAGGGA TCTCTTAGTG CTCATTACAC CCACTGCAA GAGCCCCAGG TTTTACTGCA	1620
TCATCAAGTT GCTGAAGGGT CCAGGCTTAA TGCTGGCTCT TTTCTGCCCT CAGGTCCCTGC	1680
CGGCTAAACT CTAAGGATAG GCCATCCCTCC TGCTGGTCA GACCTGGAGG CTCACCTGAA	1740
TTGGAGCCCC TCTGTACCTA TCTGGCAAC AAAGAACCT ACCATGAGGC TGGGGCACAA	1800
TGAGCTCCA CAACCACAGC TTTGGTCCAC ATGATGGTCA CACTTGGATA TACCCCAGTG	1860
TGGGTAAGGT TGGGTATTG CAGGGCCTCC CAACAATCTC TTTAAATAAA TAAAGGAGTT	1920
GTTCAGGTAA	1930

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 560 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCCAGGCAGC GGTGGGGGGA CAACCTCGTG TGCCACGCC G TGACGGCAG CATCCTGGCT	60
GGCTCCTGCC TCTATGTTGG CCTGCCCCCA GAGAAACCCG TCAACATCAG CTGCTGGTCC	120
AAGAACATGA AGGACTTGAC CTGCCGCTGG ACGCCAGGGG CCCACGGGG A GACCTTCCTC	180
CACACCAACT ACTCCCTCAA GTACAACCTT AGGTGGTATG GCCAGGACAA CACATGTGAG	240
GAGTACCAACA CAGTGGGGCC CCACTCCTGC CACATCCCCA AGGACCTGGC TCTCTTACG	300
CCCTATGAGA TCTGGGTGGA GGCCACCAAC CGCCTGGCT CTGCCCCCTC CGATGTACTC	360
ACGCTGGATA TCCTGGATGT GGTGACCAAG GACCCCCCGC CCGACGTGCA CGTGAGCCGC	420
GTCGGGGGCC TGGAGGACCA GCTGAGCGTG CGCTGGGTGT CGCCACCCGC CCTCAAGGAT	480
TTCCTTTTC AAGCCAAATA CCAGATCCGC TACCGAGTGG AGGACAGTGT GGAATGGAAG	540
GTGGTGGACG ATGTGACCAA	560

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1391 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 1..1053

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ACC CTC AAC GGG CGC CGC CTG CCC CCT GAG CTC TCC CGT GTA CTC AAC	48
Thr Leu Asn Gly Arg Arg Leu Pro Pro Glu Leu Ser Arg Val Leu Asn	
1 5 10 15	
GCC TCC ACC TTG GCT CTG GCC CTG GCC AAC CTC AAT GGG TCC AGG CAG	96
Ala Ser Thr Leu Ala Leu Ala Asn Leu Asn Gly Ser Arg Gln	
20 25 30	
CGG TCG GGG GAC AAC CTC GTG TGC CAC GCC CGT GAC GGC AGC ATC CTG	144
Arg Ser Gly Asp Asn Leu Val Cys His Ala Arg Asp Gly Ser Ile Leu	
35 40 45	
GCT GGC TCC TGC CTC TAT GTT GGC CTG CCC CCA GAG AAA CCC GTC AAC	192
Ala Gly Ser Cys Leu Tyr Val Gly Leu Pro Pro Glu Lys Pro Val Asn	
50 55 60	
ATC AGC TGC TGG TCC AAG AAC ATG AAG GAC TTG ACC TGC CGC TGG ACG	240
Ile Ser Cys Trp Ser Lys Asn Met Lys Asp Leu Thr Cys Arg Trp Thr	
65 70 75 80	
CCA GGG GCC CAC GGG GAG ACC TTC CTC CAC ACC AAC TAC TCC CTC AAG	288
Pro Gly Ala His Gly Glu Thr Phe Leu His Thr Asn Tyr Ser Leu Lys	
85 90 95	
TAC AAG CTT AGG TGG TAT GGC CAG GAC AAC ACA TGT GAG GAG TAC CAC	336
Tyr Lys Leu Arg Trp Tyr Gly Gln Asp Asn Thr Cys Glu Glu Tyr His	
100 105 110	
ACA GTG GGG CCC CAC TCC TGC CAC ATC CCC AAG GAC CTG GCT CTC TTT	384
Thr Val Gly Pro His Ser Cys His Ile Pro Lys Asp Leu Ala Leu Phe	
115 120 125	
ACG CCC TAT GAG ATC TGG GTG GAG GCC ACC AAC CGC CTG GGC TCT GCC	432
Thr Pro Tyr Glu Ile Trp Val Glu Ala Thr Asn Arg Leu Gly Ser Ala	
130 135 140	
CGC TCC GAT GTA CTC ACG CTG GAT ATC CTG GAT GTG GTG ACC ACG GAC	480
Arg Ser Asp Val Leu Thr Leu Asp Ile Leu Asp Val Val Thr Thr Asp	
145 150 155 160	
CCC CCC GAC GTG CAC GTG ACC CGC GTC GGG GGC CTG GAG GAC CAG	528
Pro Pro Pro Asp Val His Val Ser Arg Val Gly Gly Leu Glu Asp Gln	
165 170 175	
CTG AGC GTG CGC TGG GTG TCG CCA CCC GCC CTC AAG GAT TTC CTC TTT	576
Leu Ser Val Arg Trp Val Ser Pro Pro Ala Leu Lys Asp Phe Leu Phe	
180 185 190	
CAA GCC AAA TAC CAG ATC CGC TAC CGA GTG GAG GAC AGT GTG GAC TGG	624
Gln Ala Lys Tyr Gln Ile Arg Tyr Arg Val Glu Asp Ser Val Asp Trp	
195 200 205	
AAG GTG GTG GAC GAT GTG AGC AAC CAG ACC TCC TGC CGC CTG GCC GGC	672
Lys Val Val Asp Asp Val Ser Asn Gln Thr Ser Cys Arg Leu Ala Gly	
210 215 220	
CTG AAA CCC GGC ACC GTG TAC TTC GTG CAA GTG CGC TGC AAC CCC TTT	720
Leu Lys Pro Gly Thr Val Tyr Phe Val Gln Val Arg Cys Asn Pro Phe	
225 230 235 240	
GGC ATC TAT GGC TCC AAG AAA GCC GGG ATC TGG AGT GAG TGG AGC CAC	768
Gly Ile Tyr Gly Ser Lys Lys Ala Gly Ile Trp Ser Glu Trp Ser His	
245 250 255	
CCC ACA GCC CCC TCC ACT CCC CGC AGT GAG CGC CCG GGC CCG GGC GGC	816
Pro Thr Ala Ala Ser Thr Pro Arg Ser Glu Arg Pro Gly Pro Gly Gly	
260 265 270	
GGG GCG TGC GAA CCG CGG GGC GGA GAG CCG AGC TCG GGG CCG GTG CGG	864
Gly Ala Cys Glu Pro Arg Gly Gly Glu Pro Ser Ser Gly Pro Val Arg	
275 280 285	

- 84 -

CGC GAG CTC AAG CAG TTC CTG GGC TGG CTC AAG AAG CAC GCG TAC TGC Arg Glu Leu Lys Gln Phe Leu Gly Trp Leu Lys Lys His Ala Tyr Cys 290 295 300	912
TCC AAC CTC AGC TTC CGC CTC TAC GAC CAG TGG CGA GCC TGG ATG CAG Ser Asn Leu Ser Phe Arg Leu Tyr Asp Gln Trp Arg Ala Trp Met Gln 305 310 315 320	960
AAG TCG CAC AAG ACC CGC AAC CAG CAC AGG ACG AGG GGA TCC TGC CCT Lys Ser His Lys Thr Arg Asn Gln His Arg Thr Arg Gly Ser Cys Pro 325 330 335	1008
CGG GCA GAC GGG GCA CGG CGA GAG GTC CTG CCA GAT AAG CTG TAGGGGCTCA Arg Ala Asp Gly Ala Arg Arg Glu Val Leu Pro Asp Lys Leu 340 345 350	1060
GGCCACCCCTC CCTGCCACGT GGAGACCGAG AGGCCGAACC CAAACTGGGG CCACCTCTGT ACCCCTCACTT CAGGGCACCT GAGCCCTCA GCAGGAGCTG GGGTGGCCCC TGAGCTCCAA CGGCCATAAAC AGCTCTGACT CCCACGTGAG GCCACCTTG GGTGCACCCC AGTGGGTGTG TGTGTGTGTG TGAGGGTTGG TTGAGTTGCC TAGAACCCCT GCCAGGGCTG GGGGTGAGAA GGGGAGTCAT TACTCCCCAT TACCTAGGGC CCCTCCAAAAA GAGTCCTTTT AAATAAATGA GCTATTTAGG TGCAAAAAAA AAAAAAAA A	1120 1180 1240 1300 1360 1391

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 350 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Thr Leu Asn Gly Arg Arg Leu Pro Pro Glu Leu Ser Arg Val Leu Asn
1 5 10 15

Ala Ser Thr Leu Ala Leu Ala Asn Leu Asn Gly Ser Arg Gln
20 25 30

Arg Ser Gly Asp Asn Leu Val Cys His Ala Arg Asp Gly Ser Ile Leu
35 40 45

Ala Gly Ser Cys Leu Tyr Val Gly Leu Pro Pro Glu Lys Pro Val Asn
50 55 60

Ile Ser Cys Trp Ser Lys Asn Met Lys Asp Leu Thr Cys Arg Trp Thr
65 70 75 80

Pro Gly Ala His Gly Glu Thr Phe Leu His Thr Asn Tyr Ser Leu Lys
85 90 95

Tyr Lys Leu Arg Trp Tyr Gly Gln Asp Asn Thr Cys Glu Glu Tyr His
100 105 110

Thr Val Gly Pro His Ser Cys His Ile Pro Lys Asp Leu Ala Leu Phe
115 120 125

Thr Pro Tyr Glu Ile Trp Val Glu Ala Thr Asn Arg Leu Gly Ser Ala
130 135 140

Arg Ser Asp Val Leu Thr Leu Asp Ile Leu Asp Val Val Thr Thr Asp
145 150 155 160

Pro Pro Pro Asp Val His Val Ser Arg Val Gly Gly Leu Glu Asp Gln
165 170 175

Leu Ser Val Arg Trp Val Ser Pro Pro Ala Leu Lys Asp Phe Leu Phe

- 85 -

180	185	190
Gln Ala Lys Tyr Gln Ile Arg Tyr Arg Val Glu Asp Ser Val Asp Trp		
195	200	205
Lys Val Val Asp Asp Val Ser Asn Gln Thr Ser Cys Arg Leu Ala Gly		
210	215	220
Leu Lys Pro Gly Thr Val Tyr Phe Val Gln Val Arg Cys Asn Pro Phe		
225	230	235
Gly Ile Tyr Gly Ser Lys Lys Ala Gly Ile Trp Ser Glu Trp Ser His		
245	250	255
Pro Thr Ala Ala Ser Thr Pro Arg Ser Glu Arg Pro Gly Pro Gly Gly		
260	265	270
Gly Ala Cys Glu Pro Arg Gly Gly Glu Pro Ser Ser Gly Pro Val Arg		
275	280	285
Arg Glu Leu Lys Gln Phe Leu Gly Trp Leu Lys Lys His Ala Tyr Cys		
290	295	300
Ser Asn Leu Ser Phe Arg Leu Tyr Asp Gln Trp Arg Ala Trp Met Gln		
305	310	315
Lys Ser His Lys Thr Arg Asn Gln His Arg Thr Arg Gly Ser Cys Pro		
325	330	335
Arg Ala Asp Gly Ala Arg Arg Glu Val Leu Pro Asp Lys Leu		
340	345	350

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TCCAGGCAGC GGTCGGGGGA CAAC

24

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TTGCTCACAT CGTCCACCAC CTTC

24

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6663 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CCCAGAACTC TTGGACGCTG AGGCAGGAGG ATTCCCAAGT TTCAAGACAG TGTGTTCTA	60
GGTAATGAGA CCCTGTCAAG AAAAGAAAAG AAATAAAGAG ACAAGAAAAT GTTTATAGC	120
TGTGAGACAG CTTGGTGGGT AAGGGCACT TGCCCTCCAAT CAAGATGACC TCAGCCCCAT	180
CCCTAGGAAT CCATGOTAGA AGGAGAAAAGC AAACTCGCAG CTGCTGACCT CCATACATGT	240
GCTCCAATGT GCACACACAC AGGGAGACAT AATCAATTAA TAGGATGTAT TTGCTTAGAT	300
TTGAGTAGGC ATTTATGACT GATGTTTAA AATTTTTATT TGATTTATG AAAATATACC	360
TGTTTGTTATT TGGTTTGGTT TGGTTTGGT TTTGTTTATT TGAGACAGGG CTTCTCTGTG	420
TAGTCCTGGC TGTCTTGGG ACTCACTCTG TAGACCAGGC TGGCCTTGAA CTCAGAAATC	480
CGCCTGCTTG TGCTTCCCAA GTGCTTAGAT TAAAGGTGTG CACTGCCATT CAGCAAATT	540
GCATACTTTA ACCCCCAGTAT TTGGGAGGCA GAGGCAGACT AATGTGTGAA TTCCAGGCTA	600
GCCAAGGATA CAGAGTGAGA CCCTATTCTT ACCCTCCCCC CCCAAAACCC CAAAATGTAT	660
TTTGTGCTTG TGTATGTACA TGTGTTGTC AGCACGTAAA TGTCCAAGGA CAACTTGTAG	720
AAGTTCTCTC CGTTCACAGT CTAAGTCTG AATTCAAACCT AAGGTCCCTCA GGCTTAGCCA	780
CAGTCTTCTT TATGTACTGA GCCATTCAAC TGGCCCTGGG TTGACTGATG AATTAATT	840
TGAGATAAGG TCTCTTGAG CTCTAGCTAG GCTCAAACCA TGAACCTCCCAGG CAGCACTTCT	900
GAGCTGCTGG TACTCTTGCT TCCACCCCAA GTGGTGGAAAT GATACTCAGG CAGCACTTCT	960
CTGGGGAGG GGCTGGCCCTT GGCTTGATT TTGTTGCCTC ACCTCAATG AGTGTGTTGGG	1020
TCTCGTTGTT TCTTTCTTT ATCTGTGAAA TGGGTGAACA CCTGTTCAAG ACTTCCTGAC	1080
TCTTGAAACA TCCAGGCAGG GTGAGGGACT TGAAGTGGGC TCATCCCATG CCTAACAAAG	1140
TGTGCTCTTT GACCCCCAGAC ACAGCTGTAA TCAGCCCCCA GGACCCCCACC CTTCTCATCG	1200
GCTCCTCCCT GCAAGCTACC TGCTCTATAC ATGGAGACAC ACCTGGGGCC ACCGCTGAGG	1260
GGCTCTACTG GACCTTCAAT GGTCGCCGCC TGGCCCTCTGA GCTGTCCCAGC CTCCCTAAC	1320
CCTCCACCCCT GGCCCTGGCC CTGGCTAACCC TTAATGGGTC CAGGCAGCAG TCAGGAGACA	1380
ATCTGGTGTG TCACGCCCCA GACGGCAGCA TTCTGGCTGG CTCCCTGCCTC TATGTTGGCT	1440
CTAAAGTGGGG CCCCAGACAC TCAGAGATAG ATGGGGGTTG CCAATGACAG ATTTAGAGCC	1500
TGGGTCTTCT GTCCTGGGGC AGAGCCATGG GCTCTCACTT GCATGCAGGC ATGGTCATAC	1560
CCAGCACAGG CATTGCAACT CTAGGGACAG CTGTTGCTGC ACTGTCCCCCT GTGTACCCCA	1620
CAGCTTTAGA AAAGCTGTCA TGTGTTCTT GTAGTGGCCC CTGAGAAGCC CTTTAACATC	1680
ACCTGCTGGT CCCGGAACAT GAAGGATCTC ACCTGGCCCT GGACACCGGG TGCACACGGG	1740
GAGACATTCT TACATACCAA CTACTCCCTC AAGTACAACC TGAGGTTGGT ACCCAGCCAA	1800
CCCTTGCTGT GTGACTTCTG GCAATACTTA CCTTCTCTGA TCAAATATGT TCCCTGTTAT	1860
GAACATCAAAA GGGACTCTCG CACCTCCACA GGTGGTACGG TCAAGGATAAC ACATGTGAGG	1920
AGTACCAACAC TGTGGGCCCT CACTCATGCC ATATCCCCAA GGACCTGGCC CTCTTCACTC	1980
CCTATGAGAT CTGGGTGGAA GCCACCAATC GCCTAGGCTC AGCAAGATCT GATGTCCCTCA	2040

CACTGGATGT CCTGGACGTG GGTGAGCCCC CAGTGTCCAC CTGTGTTCTG CCCTAGACCT	2100
TATAGGGCGC CTCCCCCCC TCCCCCCCAGA CTTTTGGTT CTTCTAGAGG TCTTAGCCAC	2160
AGCCACGGTG GTTGCAGGAC AGTGGTTGTT CATAACTTAA TGCAAAGACT TTCCCCCAAG	2220
ACAGTCAGA TTTTCCCCCT CCCCACCCCC AACACACACA TACACACACAA CTCTGCAGAG	2280
AACACCTGGC CTGACCACCC TCCCTCTCTA CAGCCCAGGT GTTCAGAAGG GAGTCCTAGG	2340
GGACTGAGAG GAGGCCGCCCC GGTCTGAAGG CGCCCCAGGA AGCCGAGGCC TTGAGCTGGG	2400
GGGGGGGGCG AGGGTTGGAG GCACGAAC TGATGATCCCT GAGCACAACT GGGCTTAATC	2460
TAATTAGGGT GTTCCCAGCC CAAAGCAGCC TGGCCATT AACCCTCAA GTGCCTCACT	2520
GAAGACTCAG GGGAGAGATC AGCTTGTACT CTCTCCATGG TCCCCCAGGA GGGTCCCTGG	2580
GTGCCCCCTGG CTCATTCCCCA CATCCAGAGG TTTTGTGTCT TCCTGGCATC TAACCCTCAG	2640
TTGTGCTCTG TGGCTGGCAC AGCTGCCCCG TGGAGGCTCT TGGTAATGTA CAAGGCATCA	2700
GAGGTGGACA TGGGATGGGG ATACATACGG ATGGAGCCAA ATAGCACCTC AAGGTGGGT	2760
GATATACAAT AAAGCTTGTG ACCCTGACGC TCAGAAAGCC TACTCATGAT GATCACAATT	2820
GTTGACATCA CTCTGGACA TGTAGTGAGA CCCTAGCTCA AAACACAGAC AGTAGCTTA	2880
AGAGTCAGCT TGTGACTTAA TACTGGAAC CAGGGCCTAA TAGGTGCTGG GTGATGCTCG	2940
CCTCACTCCC TGTAGTGTGA GATCTCTCG CTAATCTCCA CCCCAGCTGG GTGGGCTGCT	3000
CTGTCCTCTT GAGGGCAGGA ATGTGTGTCT TCCATCAGAG ATAGGACCCG TGGTAGCAGC	3060
AACTGCTGCT GGCTGTTTCT CGAACATTTAA ATGACAGTAA TCTATCAGGC CTGGGTGAGT	3120
AGCTAACAGG GGTGGGGGGG TGGTCTGGAA AACGGAGATA CGGTCACTAGG AGCCACTGCA	3180
GCCTAGATTA CACCACTGGG TGTCTGTCA CTAGGCCATT CTCACCAAGC AGTCCTCAGA	3240
ACTGGGAGCA CTGTTGCCAG CATTAAATGC CAGCAATTAA TGCCAGCATT AGGGGAGGCA	3300
GAGGCCAGAAG GATCTCTCTG AGTTCAAGGC CATCCTGAAT TTACATAAAG AGCTCCAGGC	3360
CAGCCAGGGT GCGCAGTAAA ACCTTGTCTC AAAAACAAA GCATCTTCTAG TGACCCAGGCT	3420
TGCTCCACCC CCAGTGACCA CGGACCCCCC ACCCGACGTG CACGTGAGCC CGGTTGGGGG	3480
CCTGGAGGAC CAGCTGAGTG TGGCCTGGGT CTCACCCACCA GCTCTCAAGG ATTCCTCTT	3540
CCAAGCCAAG TACCAAGATCC GCTACCGCGT GGAGGACAGC GTGGACTGGA AGGTGCCCGT	3600
CCCGCCCCGG ACCCGCCCCCT GACCCCGCCC CCCGCATCTG ACTCCTCCCT CACCGTGCAG	3660
GTGGTGGATG ACGTCAGCAA CCAGACCTCC TGCGCTCTCG CGGGCCTGAA GCCCGGCACC	3720
GTTTACTTCG TCCAAGTGCG TTGTAACCCA TTCGGGATCT ATGGGTGAA AAAGGCGGGA	3780
ATCTGGAGCG ASTGGAGCCA CCCCCACCGCT GCCTCCACCC CTCGAAGTGG TGACCCACCTC	3840
TCCAGGGCTG GCTGGCCCAT GGAATCCCCA ATCCATCCTG TTCCCTCCCC CCCACCCCTT	3900
TTTGAGACA GCGTCTTCAG GTAGCGCATG CTGGCCTTAA ATTCAAGTATG TAGTCAAGGA	3960
TGACCTCGAG CTCCCTGGTCT TTTTGTCTCC ACTTAGAGAC AATGGCCAGT GGCCATCACC	4020
ACCTTGGGA GACTAGCCAT GGAGTCTATT TAGCCTGTCA TTTGGTGACA GATGGAGTAC	4080
AAACAGTGTGA CCTCTGTAA GAGAACTGAA GACAGGCTGT TTTTAACCCC AATATCCTAG	4140
GCTCTCTAGA GGTAAACTTT ATATAAAATA GAGACTATTA CAGCCAGTTA TCACATGGTC	4200
CCACAGAACCC TTTTGTCA CAACCTATAG ACCACAGTGC CTGTGCCTAC CACATAAGGG	4260

TCTCTACTGC TGGCCCACCC CTCCAACCCT TAAAAGGTAAC CCTAGGCAGC CTTAATATTT	4320
GCAATCCTCC TACCTCAGCC TCTTGAATGC TCAGAAACCA GGCAATTAAACC CAAGMTTCTC	4380
TTCTCTGGGT CCCCTTCCTTA AGGTGGGAGG GCCTAAAGAT GACTTCCTTT GTCCCTGAAGA	4440
CTCTCCGAGC CCATGGATCT GCACCTCTCA ATATGAATA TATTGCATAA AATGTCTGGC	4500
CTCAGTTTCC CCACCTGTCA GGTTTAGGCA GCACAGTCGG TCCAAGACAC TTCATTATTT	4560
GCAGGCAGTA TAAGAAGAACG CTCCCACCCC CCACCCGCTT CCTCCGGTCC CTAAGACAGA	4620
ATACTTCTAC ACTGAAACTG AACTCTCGCA GACGCATATG CTCACCTTAA TGATGATGAA	4680
ATAATGGGAA AACTGAGGCT CCGAGAGATT CCTGGAGGAA GAGGGTCAAA ACCAGCTCCA	4740
GGAAGCTCTC CAGCCCCCAT CCGGGCCTCT CCAGGTTCTG GGCTTGGCGG GAGTGAAACAC	4800
AGCTGGGAGG GGCTGGAGCC TGGGACCTTT GGCCCTGCT CGTGCCCAGC ACCTGCCATT	4860
CTTGCACGGG AGCCAGCAGG CGGCTGCCTC CGCCCCAGAG ACTGAAGAACG CCGGGGGTAG	4920
GGTTGGAGGG AGGTAAGCAG GGGCTCTGGG GGCCGAAGCT TGTGCCAGGG CCTGTCAGCG	4980
AGTCCCCAGT TTTATTTATG GCGTGAGGCC GATGTCCTTA TCCGCTGGCC TGCTGGGGA	5040
TGGCTGCGGC TGGGATTGG ACCCAAGGGC TGGCTTCCCA CTCAGTCCTC CAGCCCCACTC	5100
CATGTCACAC CCGTGCATTCTC TCTGAGGCTT ATCTGGAA CCCGCCCTTG TTCTGTGCTG	5160
TCTGTCCTCTA TTTCTGTCTA TCACCTTCCTC AGAGCCTTTT TTTTATGCTT TTAATATAAC	5220
TACGTTTAA AAATTGCTTT TGTATAATGT GTGTCCTTC GTGAGCGTGC GTGCCACAAC	5280
ACACACGTGA AGGTTAGAGA ACTTTGTGA GTAGGCTCTT TCCACCATGT GGGACTAGGG	5340
CTGGCGACAA GAGCAATTAC TGAGTCATCT CGCCAGCCCC TCACCCCTCA CTTCCCATCC	5400
TGTTTGATA GTCATAGGTA ATCGAAGGTA AATCGCTGGC TTTAATTTCG TAGCTATCCT	5460
GCCTCAGCCT ACCAAGTGCT GTGCTACCACT GTTTGCGGA GGGGCTCTCC TCCCAGTGTC	5520
TGGGGGTGAC ACAGTCCCAA GATCTCTGCT TTCTAGGTCT TTGTCTTAGT TTGCCCTTG	5580
CTTTGTCCGT GTCCCCTAGAG TCTCCGGCCC CACTTATCCA TTGACTGGTC TTTCCCTTAC	5640
CGAATACTCG GTTTTACCTC CCACTGATTT GACTCCCTCC TTGCTTGTC TCCATGCCG	5700
TGGCATTGCC ATTCCCTCTGG GTGACTCTGG GTCCACACCT GACACCTTTC CCAACTTTCC	5760
CCAGCCGAAG CTGGTCTGGT ATGGGAGGCC GCCGTCCCCC GCGCGCCTCC TGCTGGCCGC	5820
GCCCCAACAC TGCCGCTCCA TTCTCTTTAG AGCGCCCGGG CCCGGGGGC GGGGTGTGCC	5880
AGCCGCGGGG CGGGAGGCC AGCTCGGGCC CGGTGGGGCG CGAGCTCAAG CAGTTCTCG	5940
GCTGGCTCAA GAAGCACGCA TACTGCTCGA ACCTTAGTTT CCCGCTGTAC GACCAGTGGC	6000
GTGCTTGGAT GCAGAAAGTCA CACAAGACCC GAAACCAGGT AGGAAAGTTG GGGGAGGCTT	6060
GCGTGGGGGG TAAAGGAGCA GAGGAAGAGA GAGACCCGGG TGAGCAGCCT CCACAACACC	6120
GCACCTCTCTC TTCCAAGCAC AGGACGAGGG GATCCTGCC CCGGGCAGAC GGGGTGTGCC	6180
GAGAGGTAAG GGGGTCTGGG TGAGTGGGGC CTACAGCAGT CTAGATGAGG CCCTTTCCCC	6240
TCCCTCGGTG TTGCTCAAAG GGATCTCTTA GTGCTCATTT CACCCACTGC AAAGAGCCCC	6300
AGGTTTTACT GCATCATCAA GTTGCTGAAG GGTCCAGGCT TAATGTGGCC TCTTTCTGC	6360
CCTCAGGTCC TGCCGGCTAA ACTCTAAGGA TAGGCCATCC TCCCTGCTGGG TCAGACCTGG	6420
AGGCTCACCT GAATTGGAGC CCCTCTGTAC CATCTGGCA ACAAAAGAAC CTACCAAGAGG	6480

- 89 -

CTGGGCACAA TGAGCTCCC CAACCACAGC TTTGGTCCAC ATGATGGTCA CACTGGATA	6540
TACCCCGAGTG TGGTAGGGT TGGGGTATTG CAGGGCCTCC CAAGAGTCTC TTTAAATAAA	6600
TAAAGGAGTT GTTCAGGTCC CGATGGCCAG TGTGTTGGG GCCTATGTGC TGGGGTGGGG	6660
GGA	6663

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 186 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Asp Pro Thr Leu Leu Ile Gly Ser Ser Leu Gln Ala Thr Cys Ser Ile	
1 5 10 15	
His Gly Asp Thr Pro Gly Ala Thr Ala Glu Gly Leu Tyr Trp Thr Phe	
20 25 30	
Asn Gly Arg Arg Leu Pro Ser Glu Leu Ser Arg Leu Leu Asn Thr Ser	
35 40 45	
Thr Leu Ala Leu Ala Leu Ala Asn Leu Asn Gly Ser Arg Gln Gln Ser	
50 55 60	
Gly Asp Asn Leu Val Cys His Ala Arg Asp Gly Ser Ile Leu Ala Gly	
65 70 75 80	
Ser Cys Leu Tyr Val Gly Leu Pro Pro Glu Lys Pro Phe Asn Ile Ser	
85 90 95	
Cys Trp Ser Arg Asn Met Lys Asp Leu Thr Cys Arg Trp Thr Pro Gly	
100 105 110	
Ala His Gly Glu Thr Phe Leu His Thr Asn Tyr Ser Leu Lys Tyr Lys	
115 120 125	
Leu Arg Leu Val Arg Ser Gly * His Met * Gly Val Pro His Cys	
130 135 140	
Gly Pro Ser Leu Met Pro Tyr Pro Gln Gly Pro Gly Pro Leu His Ser	
145 150 155 160	
Leu * Asp Leu Gly Gly Ser His Gln Ser Pro Arg Leu Ser Lys Ile	
165 170 175	
* Cys Pro His Thr Gly Cys Pro Gly Arg	
180 185	

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- 90 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

AGCTTGGCGCG CCTCCCCGGGC GGATCGGGAG CCCAC

35

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AGCTACGGGT TTAGAGTTA GCCGGCAG

28

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Val Leu Ala Ser Ser Thr Thr Ser Ile His Thr Met Leu Leu	15		
1	5	10	
Leu Leu Met Leu Phe His Leu Gly Leu Gln Ala Ser Ile Ser	30		
20	25		

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Ile Lys Pro Ser Gly Arg Arg Gly Ala Ala Arg Gly Pro Ala Gly Asp Tyr Lys Asp Asp
 5 10 15 20
 Asp Asp Lys

(2) INFORMATION FOR SEQ ID NO:34:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- 91 -

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GATCTTGCCC TCGGGCAGAC GGGGTGCGGC GAGAGGTCTT GCCGGCGACT ACAAGGACGA	60
CGATGACAAG TAG	73

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 73 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AACGGGAGCC CGTCTGCCCT ACAGCCGCTCT CCAGGACGGC CGCTGATGTT CCTGCTGCTA	60
CTGTTCATCC TAG	73

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CCCACGCTTC TCATCGGATT CTCCCTG	27
-------------------------------	----

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CAGTCCACAC TGTCCCTCCAC TCGGTAG	27
--------------------------------	----

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11832 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GGGGCCGCTG	CAGTGATTAC	TCACCGCGTG	GCGCACCCCA	CCCGCGGGCC	GCTGACTGGA	60
TTTTTCCGTG	GGGGGATGTG	AAGAAGTTA	GGGAGAACTC	TTCTGCACCG	ATGGGAACTA	120
GGAATGCAGG	GTTCGGTCCC	GTTCCCCCAA	GGACACACCT	CTCCCCATAA	CCCCACTCAT	180
AAGGGCTCCC	TGCACGGCCT	CCGGGACATC	CCCATATCCA	ATACCCGCAG	ATATGATAGT	240
TGAGAAGGGA	CCAGAGGCCG	GAGACTCCCT	CCCTGCCTTC	TGGCTTCCC	CCCCCCCTGC	300
ACGAAACGAG	ACTACAGCGA	TGGGAGAGGT	GGCATGAAGG	CTTAGGGTGG	GGATCGGTAG	360
GACCCATGCA	CCCAGAGAAA	GGGACTGGTG	GCAAACCTTC	AACTCTCTGG	GGAAGGAAGA	420
AGGGCTGAAA	GAGGATGAAC	GGGCTCAGGT	ACTGCTCAAT	GTGTGTGTGG	CGGACCAAAG	480
TGGGTATGGG	GGCCCCGTAA	GAGGGCCGGG	GAAGGTGGAT	AGGAAGGATC	CCGGTAGACT	540
GGAGGGGATC	CTGGAAAAGC	ACCAGGGCTG	CGAGCTAGGA	ACCCATTCCG	AGTAAAGGGT	600
ACAGGATCCC	AGATGAGGGG	TGGGAAGCC	TGGACGGGC	GGGACCAGAG	AGGGAGGTCC	660
CACGGGCTGG	TGGGGAAAGA	GTGGGGGCT	TCGGCGAGGA	GGATGGGACG	TTCAGGAGTG	720
GTAACTGGGC	GGAGGCCGGC	CGGGCGGGC	GGCGGGTGCC	CGCGGGCGGT	GGGAAGGCCG	780
GTGCGGGGCC	CACGATCAAC	CCCCCCAG	GGGCCGGGCC	GGGCCGGGGG	CGGGGCCGGG	840
CGGGCGGAGC	GGCCGATTAG	CGCCTTGTCA	ATTCGGCTG	CTCAGACTTG	CTCCGGCCTT	900
CGCTGTCCGC	GCCCAGTGA	GCGCGTGAGG	ACCCGAGCCC	CAATCTGCAC	CCCGCAGACT	960
CGCCCCCGCC	CCATACCGGC	GTGAGTCA	CCGCCCCGTG	CGCGCCACCC	CCATGCCCGC	1020
GGGTGCGCCG	GGCCCCGTG	CCCAATCCGC	GGGGCGGCCG	CGCGGGCCGC	TGTCTCGCT	1080
GTGGTCGCCT	CTGTTGCTCT	GTGTCTCGG	GGTGCCTCGG	GGCGGATCGG	GAGCCCGTGA	1140
GTACCGTGGC	CCCTGCTCCC	CACCTCCCCA	GGGAAGCCGG	GATCCGGCCG	CCCGGGGGGT	1200
AGTCGCGGGG	GATGGAAGAA	GGGGCGCGAG	CGCCACCTGG	ACGTCCCCGG	AACAAAGGAA	1260
GGCGGCCCTC	GGGGCGCCCT	CACCTGTGGG	GCTCATGGCA	CCACCACCA	GCCTCCCAAG	1320
AGTACCCCGT	TATACATCA	AGGCCTTTA	TCTGTATCCC	CTTTGCGAGG	CTGTCTGGCC	1380
AGGCTCAGTT	TGAAGGACAT	CGCAGTGTCC	TGGGACCCCC	CTCCTTCAGG	GTGCTGGAC	1440
GCTTCGGGCC	GCACGCGCTGT	GTCTTGATA	TCAGAGCGGA	AGGGAAGCCT	CCCTGGCCGG	1500
GGGCGCACGC	TTGGGTGCGT	TGGGTTGGGT	GCTGGCGCAA	AGTGGGGTCC	CCTCCCCCAT	1560
GAAGTGATGA	TCCCCGGGGG	GAGGGTGGGG	CGTTATCGTG	AGCCCTCCTG	TCCGCGCTGGC	1620
ATGCGGCCCG	GGCTCCCTCG	GGACTTGCGT	CTCCGTGGGG	TCGGCGCCGC	CCCCCTCCCC	1680
CTATAGCAGA	CTCCATGCTT	TGGTATCCTC	GAAGTCCCTCT	CCACTGGTGG	GGCTCACAAAC	1740
CGGTCTCATT	CAGGCTGCGC	TGGGTTGAGA	GCCTCTAGCG	ACTGAAATT	CGGTGAGGAG	1800
CGAGAGCAAG	CGTGTCCGGG	CACCGCGAGC	CCAGACTTCA	TTGTCTAAGG	GGCACCCAGT	1860

GGGGGTCAGC	TGCCGAGAGA	ATCCCACGT	CCCAGGAGGA	ACTCCTGGCC	TTGAGCCCC	1920
ATCACCCAAC	GCACACATCC	CCGCCAGGAT	GCGGCTCTCCA	CATCCAGACC	CTCTCTGGGA	1980
CACACCCAAA	GACACACAAA	AGAGCCCCAC	TGGCTTATGT	CCCGTCACCC	TGCCCTCCGA	2040
CGCGCGCTGC	AGCCCAGATG	CGTATTGCAC	CACCATCGCG	GGCCTCGCAT	TCCATCCTCT	2100
ACACACACAC	ACACACACAC	ACACACACAC	ACACACACAC	ACACACAGAC	ACGCACACAC	2160
ACACGCACGC	ACACACACGC	ACGCCCGCAC	TCGTGGTCCC	ACATTTATTT	CACAGGGGAG	2220
GCAACACCGG	GGTACGCATA	TGGTTGAGTG	CACTGGAGAT	CTTCCCCAC	CACTCTCAGG	2280
ACCCCATCCG	GAGACACAGG	CCACACCGCA	GGGGCACCAC	GCTGCCTGC	TGCTCTGGC	2340
TAGTAGCTT	GTGCAGTTG	TCCGCGGTGT	CTGTGGACGC	CCTCCCGCTC	TTGTCAGGGG	2400
ACAGGAACCT	ACACTCCTGC	TTGCCCAAGG	CGGCTGGCA	GGTGTATGTGG	TGACACCCGG	2460
GACCTTCCG	GGGAGTTGGT	GTTGCTGCCA	AGCCTGGTA	GTTTTTGAAT	GCCACCAATA	2520
GCGCTAACGCT	TTGTTTCCGG	GCGGGCTGCA	GAGCAACAGG	CGAAGGTGGC	GGAGTGGGGG	2580
TGGCGCGTGT	GTTTTTCTT	TTAAGGGGGA	GAGAAATTAA	ATAAGAGGTT	CTCACACCTC	2640
TGCAATCTGT	TTGTACTTAC	CGTGTGTCTT	AACACCTGAC	CAGCCAGCCG	GTGGGTCGTA	2700
AAAGTGTATG	CAGGTACCG	CGGGACAGGA	GATGGGGGCC	CCTGGGTAT	GGCTGGGATG	2760
GAGGCCACCT	TCCCCTGGC	CTTTCAGGGA	ATCTCACACT	TTTCCCTTTT	AAAACACATG	2820
GTGTTCTTT	TAATAACGGC	AGCAAACCTCG	CATTGGAAA	GGGGAAATA	AGCTTGTATA	2880
GGCCCCGGCT	TTGTGGAAAG	GAGGGGAAGA	GGGAAGAAAA	AAGGAGGGGT	GTCTCCTCCA	2940
GGCTTAGGGG	GCTGTCAGCT	GCTGCTCTGT	CTAGCTTGGC	ATGTGTGTGC	CCCAGTCCCC	3000
AGTGGCTTTG	GCCCATTGTT	TGTGGAAGCC	AAGAGGGAGA	CTGGAGTCCT	CTATCTCTGG	3060
TACTCCAGAG	TCAGGCTTCT	CAGTCGGAGC	CCAGAGAACG	TCTTCCCTGT	TTTATGGAGG	3120
GAATCAGGGG	AGGGGGTGCC	AGGTGGACTA	CGTTCTGCTG	AGGACTGTAC	CAGTCGCTCG	3180
AAGGAGAAAG	CTTGGGCTTG	CCCCCTCTCC	CCCTCAAGCC	ACGAAGGGCA	GCTGCTAGGC	3240
TAGTGTGGTA	AAAGGGCATT	ACTCCCCAGC	CAGGACCCCC	CAGAGAGTCC	CCTTCCTGGC	3300
CAGACAAATG	CTGGGGAGGG	ACAGAGGGGT	GTGATCATTG	CCCAGGAGTG	CAGACAGTGG	3360
GGTCCCCGGT	CGGGCAGTGC	CTCCCACCCCT	GCTGAGGGGG	GGCCCCAGGC	AGGAAGCCGT	3420
GGGTGGGCCG	GGGTAGAGAC	GCTGGCACGT	CCCAGTTCAT	GCCGAAGGAA	TTCTGAATTA	3480
GGGGGGGGCT	GGCTGCCTGG	GACCTCCGGG	GGGGCCCCCT	GGCCCCCGCC	GCTCCGTCTG	3540
GCCTGCTCCT	CCTGCTCCTT	CGCACGGACG	CTGAGACCTC	CGCTGAGCCC	TGGGACAAGC	3600
CCCAAAATGCA	ACTGCGATTG	CAGGCTTCGC	AAGACCCGCC	TCCTCCCAAG	GCCAAATTG	3660
CCTGGGAGAA	GTCATTTCAGG	GCCCAGACTA	GAACCATGTT	GGTGCACCT	CATCCATCTG	3720
GGGCATGAAG	GACCGTCCAG	GGCTGCAGTT	TAGCTCTTA	ATACGAACCT	GGGGTGGGT	3780
GCAGCCTCTG	TTCTCCGAGC	CTCTTTGGAA	ATCGGTTTG	TTTTGTTTT	TGTTTTTCC	3840
AATACTCTT	TCCCTCTCATC	CCATCCCCGG	ACTGTTTCC	TCCCTAAGGG	TTGAGAGCCC	3900
TGCAGTCTTC	CCTAACCTTT	TCTTGTCTTC	TACCCCAAGG	CCTTGCACCA	TGGAGTCCC	3960
CCTCTCCCT	TGCCCAACTG	GGGCTCCAGC	CTTACTGCAT	TTGGCTCTTG	GTAACGTCC	4020
CAGGGCCTCT	CTGACACACA	GGGTTGTAGC	CCCAGCTCCC	TCTCTCTCTCC	TCCCCCCTTT	4080

CTCTTTGCT TCTGAGACTT AATTTTTTC TTTTCTTTT TGGCTTTTG AGACAGGGTT	4140
TCTCTGTACA GCCCTGGCTG CCCTGGCACT CATTCCTGTAG ACCAGGCTAG CCTCAGACTC	4200
ACAAACCTAC CTGCCTCTGC CTTTCCAGTG CTGGCACTAA AGATGTGGGC CACCACAACT	4260
AGTAGTTAAG TGTTTGCTG TGTCTTATT CCTATAGTGA CCTCAGTTCC TGGCATATTG	4320
TAGGCATGG ATGGATGAAT GGATGGATGG ATGGATGGAT GGATGGTTGG ATGGAGCAAG	4380
CTTGAATCGT CCTGAGTGAA AAAAGAGACC TCAGAGAACT GAATGGAGTT AGGTTCCCAG	4440
GGCAGCCTGG CCTGCTGGTC TCATGGGAGC TCCCTGTGAA ACTTCCCCCA CACCTCCAC	4500
CACCCCTGCCA TCCCTGTGTGG CTGACAAGAA AGGCCAATGG CCAGATGGGG ACACAGACTC	4560
AGGAAAGCTT GGAATATGTT CCCCTCCTCA TATCCTAGGC CTTGTTGTCC CCCTGAGGGC	4620
CCAGCCTATG AGTAGGGCAG CTGCTGGCTG CCCTAACGTT GGGTAGGCAA GAAGGGGGTG	4680
GTCCTCAGG GTGGGTCACA GGATTGAGGT CATTTCAAA GTGGCCATCA CAGTGGCCCT	4740
AGGAAATGAT TGTGGAGAGT CAGAACTCCT GTTGGGAGTT GTAGAGGGCC TTGCATGTGG	4800
GCTTCTGTGG CTGTCCTTC TCTTGTGGTC CTTTGCACAG TCCCCCTCGT TGTGCTGGGA	4860
TGTGAGGAGG GCACGGGAA AATGAAGGCT CACCCCTCA GCTTGCCTT CACGGTTCAC	4920
CCAACAGGGC TCACCTCTCC TCTGGACAGG CTCTCACTGT ATGCACAGAT TGGCCTCAC	4980
TTTGATTCCC TTCCCTTGGT CTCCCTGGAT GACAAACATT TACCAAGGTA GGATTTACA	5040
TTTTAGATAT GTCCATTCTC CAGAAACACA CTTGTGAGGT TAGGGTATCA GTGAAAGGAC	5100
ACCACCAAGGA CAGACAAAGA ATTGGAGAGG AAGGAAATTG GTAAGCCAGG CCATGCTGA	5160
TGGCTTATGT GTAATCCAG AACTCTGGAC GCTGAGGCAG GAGGATTCCA AGTTTCAAGA	5220
CAGTGTGTTTC TAGGTAATGA GACCCTGTCA AGAAAAGAAA AGAAATAAAG AGACAAGAAA	5280
ATGTTTATAG GCTGTGAGAC AGCTTGGTGG GTAAGGGGCA CTTGCCTCCA ATCAAGATGA	5340
CCTCAGCCCC ATCCCTAGGA ATCCATGGTA GAAGGAGAAA GCAAACCTCA GCTGCTGACC	5400
TCCATACATG TGCTCCAATG TGCACACACA CAGGGAGACA TAATCAATTAA ATAGGATGTA	5460
TTTGCTTAGA TTTGAGTAGG CATTATGAC TGATGTTTTA AAATTTTTAT TTGATTTAT	5520
GAAAATATAC CTGTTGTAT TTGGTTTGGT TTGGTTTGAG TTTGTTTAT TTGAGACAGG	5580
GCTTCTCTGT GTAGTCTGGT CTGTCCTGG AACTCACTCT GTAGACCAGG CTGGCCTTGA	5640
ACTCAGAAAT CGCCTGCTT GTGCTTCCA AGTGCTTAGA TTAAAGGTGT GCACTGCCAT	5700
TCAGCAAAAT TGCATACTTT AACCCCAAGTA TTTGGGAGGC AGAGGCAGAC TAATGTGTGA	5760
ATTCCAGGCT AGCCAAGGAT ACAGAGTGAG ACCCTATTCT TACCCCTCCCC CCCCCAAAACC	5820
CCAAAATGTA TTTGTGCTT GTGTATGTAC ATGTGTGTTG CAGCACGTAA ATGTCCAAGG	5880
ACAACCTGTA GAAGTTCTCT CCGTTCACAG TCTAAGTCCT GAATTCAAAC TAAGGTCTC	5940
AGGCTTAGCC ACAGTCTTCT TTATGTACTG AGCCATTCA CTGGCCCTGG ATTGACTGAT	6000
GAATTAATTT TTGAGATAAG GTCTCTGTGTA GCTCTAGCTA GGCTCAAACAT ATGAACCTCCC	6060
AAGGTCTACT TGAGCTGCTG TGACTCTTGCT TTCCACCCCA AGTGGTGGAA TGATAACTCAG	6120
GCAGCAGTTC TCTGGGAAAG GGGCTGGCT TGGCTTGAT TTGTTGCCT CAGCTTCAAT	6180
GAGTGCTTGG GTCTCGTTGT TTCTTTCTT TATCTGTGAA ATGGGTGAAC ACCTGTTCAA	6240
GACTTCCTGA CTCTTGAAAC ATCCAGGCAG GGTGAGGGAC TTGAAGTGGG CTCATCCCAT	6300

GCCTAACAAA GTGTCGTCTT TGACCCAGA CACAGCTGTA ATCAGCCCC AGGACCCCAC 6360
 CCTTCTCATC GGCTCCTCCC TCCAAGCTAC CTGCTCTATA CATGGAGACA CACCTGGGC 6420
 CACCGCTGAG GGGCTCTACT GGACCTCAA TGTCGCCGC CTGCCCTCTG AGCTGTCCC 6480
 CCTCCCTAAC ACCTCCACCC TGGCCCTGGC CCTGGCTAAC CTTAATGGGT CCAGGCAGCA 6540
 GTCAGGAGAC AATCTGGTGT GTCACGCCCG AGACGGCAGC ATTCTGGCTG GCTCCTGCCT 6600
 CTATGTTGGC TGTAAAGTGGG GCCCCAGACA CTCAGAGATA GATGGGGGTT GGCAATGACA 6660
 GATTTAGAGC CTGGGTCTTC TGTCCTGGG CAGAGCCATG GGCTCTCACT TGCATGCAGG 6720
 CATGGTCATA CCCAGCACAG GCATTGCAAC TCTAGGGACA GCTGTGGCTG CACTGTCCCC 6780
 TGTGTACCCC ACAGCTTAG AAAAGCTGTC ATGTTTCCT TGTAGTGCCT CCTGAGAAC 6840
 CCTTTAACAT CACCTGCTGG TCCCCGAACA TGAAGGATCT CACGTGCCGC TGGACACCGG 6900
 GTGCACACGG GGAGACATTC TTACATACCA ACTACTCCCT CAAGTACAAG CTGAGGTTGG 6960
 TACCCAGCCA AGCCTGCTG TGTGACTTCT GGCAATACTT ACCTTCTCTG ATCAAATATG 7020
 TTCCCTGTTA TGAACTCAAAGGGACTCTC GCACCTCCAC AGGTGGTACG GTCAGGATAA 7080
 CACATGTGAG GAGTACCACA CTGTGGGCC TCACTCATGC CATATCCCCA AGGACCTGGC 7140
 CCTCTTCACT CCCTATGAGA TCTGGGTGGA AGCCACCAAT CGCCTAGGCT CAGCAAGATC 7200
 TGATGTCCTC ACACTGGATG TCCCTGGACGT GGGTGAGCCC CGAGTGTCCA CCTGTGTTCT 7260
 GCCCTAGACC TTATAGGGCG CCTCCCCCCC ATCCCCCAG ACTTTTTGGT TCTTCTAGAG 7320
 GTCTTAGCCA CAGCCACGGT GGTTGCAGGA CAGTGGTTGT TCATAACTTA ATGCAAAGAC 7380
 TTTCCCCAA GACAGTCAAG ATTTTCCCT CCCCCACCCCC AACACACACA TACACACACA 7440
 CTCTGCAGAG AACACCTGGC CTGACCACCC TCCCTCTCTA CAGCCCAGGT GTTCAGAAGG 7500
 GAGTCCTAGG GGACTGAGAG GAGGCGCCCA GGTCTGAAGG CGCCCCAGGA AGCCGAGGCC 7560
 TTGAGCTGGG GGGGGGGCG AGGGTTGGAG GCACCAACTG GATGATCCCT GAGCACAACT 7620
 GGGCTTAATC TAATTAGGTT GTTCCCAGCC CAAAGCAGCC TGGCCATT AACCCTTCAA 7680
 GTGCCTCACT GAAGACTCAG GGGAGAGATC AGCTTGTACT CTCTCCATGG TCCCCCAGGA 7740
 GGGTTCTGG GTGCCCTGG CTCATTCCA CATCCAGAG TTTTGTGTCT TCCTGGCATC 7800
 TAACCCTCAG TTGTGCTCTG TGGCTGGCAC AGCTGCCCCG TGGAGGCTCT TGGTAATGTA 7860
 CAAGGCATCA GAGGTGGACA TCGGATGGGG ATACATAGGG ATGGAGCCAA ATAGCACCTC 7920
 AAGGTGGGT GATATACAAT AAAGCTGTC ACCCTGACGC TCAGAAAGCC TACTCATGAT 7980
 GATCACAATT GTTGACATCA CTCTGGACA TGTACTGAGA CCCTAGCTCA AAACACAGAC 8040
 AGTAGTTTA AGAGTCAGCT TGTGACTTAA TACTGGAAC TAGGGCCTAA TAGGTGCTGG 8100
 GTGATGCTCG CCTCACTCCC TGTTTAGTGA GATCTCTGGC CTAATCTCCA CCCCAGCTGG 8160
 GTGGGCTGCT CTGTCCCCCTT GAGGGCAGGA ATGTGTGTCT TCCATCAGAG ATAGGACCCG 8220
 TGGTAGCAGC AACTGCTGCT GGCTGTTCT GGAATATTAA ATGACAGTAA TCTATCAGGC 8280
 CTGGGTGAGT AGCTAACAGG GGTGGGGCG TGGTCTGGAA AACGCAGATA GGGTCATAGG 8340
 AGCCACTGCA GCCTAGATTA CACCACTGGG TGTTCTGTCA CTAGGCCATT CTCACCAAGC 8400
 AGTCCTCAGA ACTGGGAGCA CTGTTGCCAG CATTAAATGC CAGCATTAA TGCCAGCATT 8460
 AGGGGAGGCA GAGGCAGAAG GATCTCTCTG AGTTCAAGGC CTCCTGAAT TTACATAAAG 8520

8607460 "✓592Z060

AGCTCCAGGC CAGCCAGGGT GCGCAGTAAA ACCTTGTCTC AAAAAACAAA GCATTTAG	8580
TGACCAGGCT TGCTCCACCC CCAGTGACCA CGGACCCCCC ACCCGACGTG CACGTGAGCC	8640
GCGTTGGGGG CCTGGAGGAC CAGCTGAGTG TGCCTGGGT CTCACCACCA GCTCTCAAGG	8700
ATTTCCCTTT CCAAGCCAAG TACCAAGATCC GCTACCCGCGT GGAGGACAGC GTGGACTGGA	8760
AGGTGCCCCGT CCCGCCCGG ACCCGCCCT GACCCCGCCC CCCGCATCTG ACTCCTCCCT	8820
CACCGTGCAG GTGGTGGATG ACGTCAGCAA CCAGACCTCC TGCCGTCTCG CGGGCCTGAA	8880
GCCCCGCACC GTTTACTTCG TCCAAGTGC G TTGTAACCCA TTCGGGATCT ATGGGTGAA	8940
AAAGGCGGGA ATCTGGAGCG AGTGGAGCCA CCCCACCGCT GCCTCCACCC CTCGAAGTGG	9000
TGACCACCTC TCCAGGGCTG GCTGGCCCAT GGAATCCCCA ATCCATCCTG TTCCCTCCCC	9060
CCCACCCCTT TTTTGAGACA GCGTCTTCAG GTAGCGCATG CTGGCCTTAA ATTCAAGTATG	9120
TAGTCAGGA TGACCTCGAG CTCCCTGGTCT TTTTGTCTCC ACTTAGAGAC AATGGCCAGT	9180
GGCCATCACC ACCTTTGGGA GACTAGCCAT GGAGTCTATT TAGCCTGTCA TTTGGTGACA	9240
GATGGAGTAC AACAGTGTGA CCTCTTGAA GAGAACTGAA GACAGGCTGT TTTAACCCCC	9300
AATATCCTAG GCTCTCTAGA GGTTAACCTT ATATAAAATA GAGACTATTAA CAGCCAGTTA	9360
TCACATGGTC CCACAGAACCC TTTTGTACA CAACCTATAG ACCACAGTGC CTGTGCCTAC	9420
CACATAAGGG TCTCTACTGC TGGCCACCC CTCCAACCCCT TAAAAGCTAA CCTAGGCAGC	9480
CTTAATATTT GCAATCCTCC TACCTCAGCC TCTTGAATGC TCAGAAACCA GGCATTAACC	9540
CAAGTTCTC TTCTCTGGGT CCCTTCTTA AGGTGGGAGG GCCTAAAGAT GACTCCTTT	9600
GTCCTGAAGA CTCTCCGAGC CCATGGATCT GCACCTCTA ATATGAAATA TATTGCATAA	9660
AATGCTGGC CTCAGTTCC CCACCTGTCA GGTTAGGCA GCACAGTCGG TCCAAGACAC	9720
TTCATTATTT GCAGGCAGTA TAAGAAGAAG CTCCCATCCC CCACCCGCTT CCTCCGGTCC	9780
CTAAGACAGA ATACTTCTAC ACTGAAAATG AACTCTCGCA GACGCATATG CTCACTTTAA	9840
TGATGATGAA ATAATGGGG AACTGAGGCT CCGAGAGATT CCTGGAGGAA GAGGGTCAA	9900
ACCAGCTCCA GGAAGCTCTC CAGCCCCAT CCGGGCCTCT CCAGGTTCTG GGCTTGGCGG	9960
GAGTGAACAC AGCTGGGAGG GGCTGGAGCC TGGGAGCTT GGCCCTTGCT CGTGCCTAC	10020
ACCTGCGATT CTTGCACGGG AGCCAGCAGG CGGCTGCCTC CGCCCGAGAG ACTGAAGAAG	10080
CCGGGGTAG GGTTGGAGGG AGGTAAAGCAG GGGCTGTGGG GGCGAAGCT TGTGCCAGGG	10140
CCTGTCAGCG AGTCCCCAGT TTTATTTATG GCGTGAGGCC GATGTCTTA TCCGCTGGCC	10200
TGCTGGGGGA TGGCTGCGGC TGGGGATTGG ACCCAAGGGC TGGCTTCCCA CTCAGTCCTC	10260
CAGCCCACTC CATGTCACAC CCGTGCATTC TCTGAGGCTT ATCTTGGAA CCCGCCCTG	10320
TTCTGTGCTG TCTGTCTCTA TTTCTGTCA TCACCTTCCC AGAGCCTTTT TTTTATGCTT	10380
TTAATATAAC TACGTTTAA AAATTGCTTT TGTATAATGT GTGTGCCTTC GTGAGCGTGC	10440
GTGCCACAAC ACACACGTGA AGGTTAGAGA ACTTTGTTGA GTAGGCTCT TCCACCATGT	10500
GGGACTAGGG CTGGCGACAA GAGCAATTAC TGAGTCATCT CGCCAGCCCC TCACCCCTCA	10560
CTTCCCATCC TGTGGATA GTCATAGGTAA ATCGAAGGTAA AATCGCTGGC TTTAATTTCG	10620
TAGCTATCCT GCCTCAGCCT ACCAAGTGCT GTGCTACCAC GTTGTGGGA GGGGCTCTCC	10680
TCCCACTGTC TGGGGTACA CAGTCCCAAG ATCTCTGCTT TCTAGGTCTT TGTCTTAGTT	10740

TGCCCCCTGCT	TTTGTCCGTG	TCCCTAGAGT	CTCCGGCCCC	ACTTAGTCTC	CATTGATITC	10800
CTTTCTGACC	GAATACTCGG	TTTACCTCC	CACTGATTG	ACTCCCTCCT	TTGCTTGCTCT	10860
CCATCGCCGT	GGCATTGCCA	TTCCTCTGGG	TGACTCTGGG	TCCACACCTG	ACACCTTCC	10920
CAACTTTCCC	CAGCCGAAGC	TGGTCTGGTA	TGGGAGGCCG	CGTCCCGCG	CGCGCCTCCT	10980
GCTGGCCGCG	CCCCAACACT	GCCGCTCCAT	TCTCTTACA	GCGCCCGGGC	CCGGGCGGGCG	11040
GGGTGTGCGA	GCCGCGGGGC	GGCGAGCCCA	GCTCGGGCCC	GGTGCAGCGC	GAGCTCAAGC	11100
AGTTCCCTCGG	CTGGCTCAAG	AAGCACCGCAT	ACTGCTCGAA	CCTTAGTTTC	CGCCTGTACG	11160
ACCAGTGGCG	TGCTTGGATG	CAGAAGTCAC	ACAAGACCCG	AAACCAGGTA	GGAAAGTTGG	11220
GGGAGGCTTG	CGTGGGGGGT	AAACGAGCAG	AGGAAGAGAG	AGACCCGGGT	GACCGACCTC	11280
CACAAACACCG	CACTCTTCAT	TCCAAGCACA	GGACGAGGGG	ATCCTGCCCT	CGGGCAGACG	11340
GGGTGCGCG	AGAGGTAAGG	GGGTCTGGGT	GAGTGGGCC	TACAGCAGTC	AGATGAGGC	11400
CTTTTCCCT	CTTCGGTGT	TGCTCAAAGG	GATCTCTTAG	TGCTCATTTTC	ACCCACTGCA	11460
AAGAGCCCCA	GGTTTTACTG	CATCATCAAG	TTGCTGAAGG	GTCCAGGCTT	AATGTGGCCT	11520
CTTTTCTGCC	CTCAGGTCC	GCCGGCTAAA	CTCTAAGGAT	AGGCCATCCT	CCTGCTGGGT	11580
CAGACCTGGA	GGCTCACCTG	AATTGGAGCC	CCTCTGTACC	ATCTGGCAA	CAAAGAAACC	11640
TACCAAGAGGC	TGGGCACAAT	GAAGCTCCAC	AACCACAGCT	TGGTCCACA	TGATGGTCAC	11700
ACTTGGATAT	ACCCCACTGT	GGGTAGGGTT	GGGGTATTG	AGGGCCTCCC	AAGAGTCTCT	11760
TTAAATAAAAT	AAAGGAGTTG	TTCAGGTCCC	GATGGCCAGT	GTGTTGGGG	CCTATGTGCT	11820
GGGGTGGGGG	GA					11832

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 26 amino acids
 - (B) TYPE: amino acids
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Val Ile Ser Pro Gln Asp Pro Thr Leu Leu Ile Gly Ser Ser Leu Gln Ala Thr Cys Ser
5 10 15 20
Ile His Gly Asp Thr Pro

- 98 -

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GTCCAAAGTGC GTTGTAAACCC A

21

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCTGAGTGTG CGCTGGGTCT CACC

24

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GGCTCCACTC GCTCCAGA

18